AN SSC PUBLICATION

Add an easy index to your CD-ROM (we did)

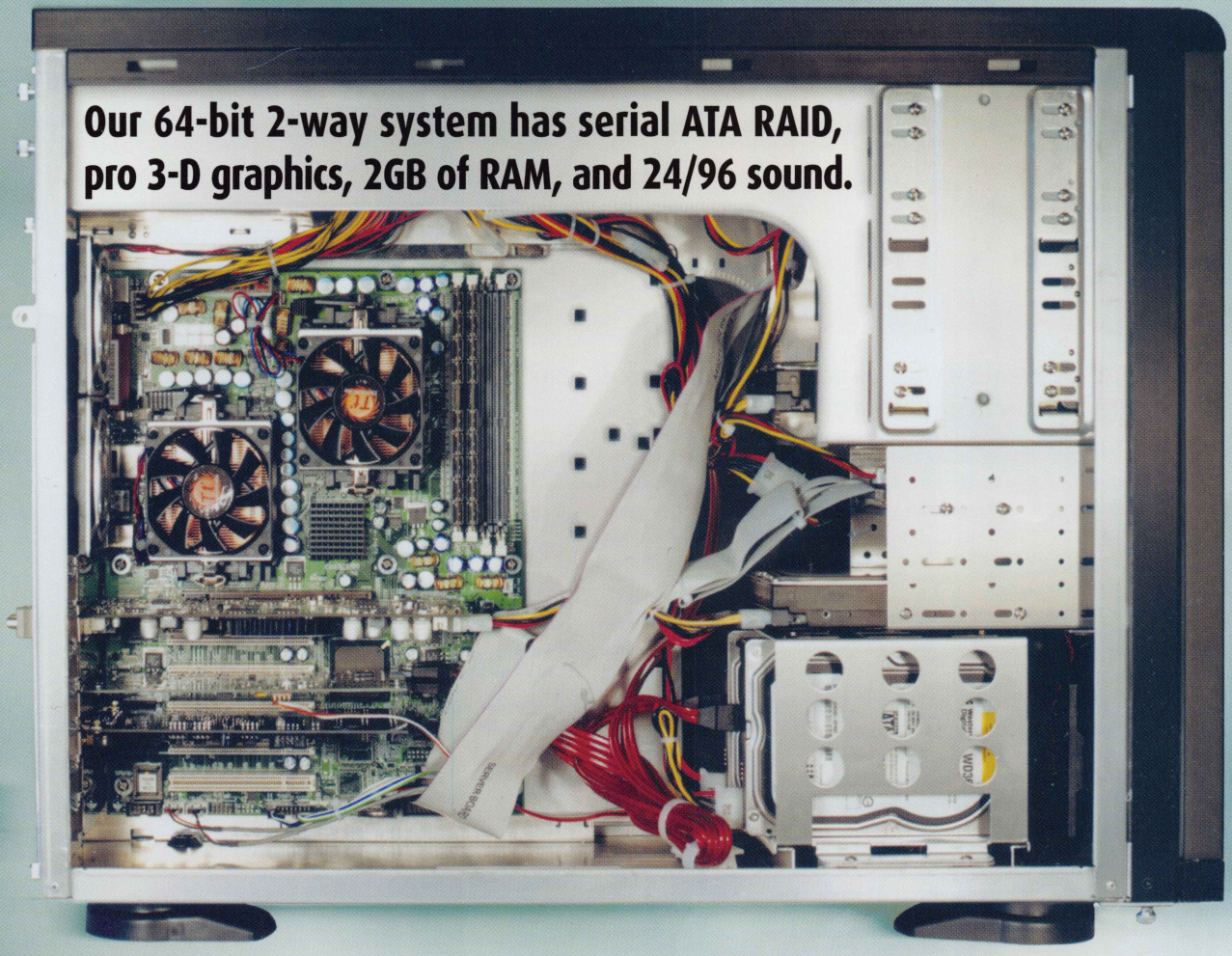
LINUX JOURNAL

We test the **Lindows laptop**

The Monthly Magazine of the Linux Community • **DECEMBER 2003**

ULTIMATE LINUX BOX 2003

Our 64-bit 2-way system has serial ATA RAID, pro 3-D graphics, 2GB of RAM, and 24/96 sound.

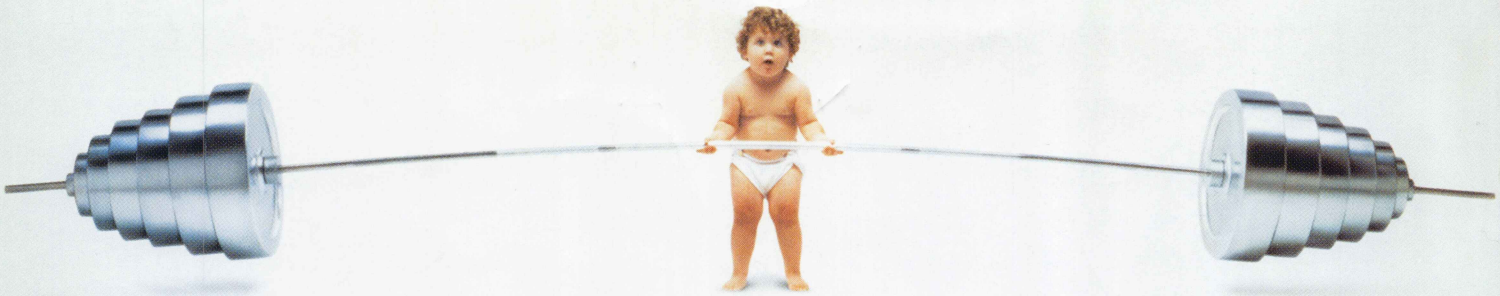


USA $5.00 CAN $6.50
www.linuxjournal.com
12
0 71486 03102 4

DVD Player Shootout | Authoring DVDs under Linux

Drag-and-drop sound design • Embedding Perl in MySQL

Kernel memory management • Secure storage with USB flash disks

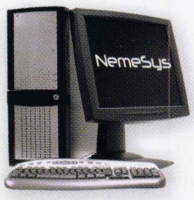


Experience the most computing power in the smallest footprint possible...

AMD Opteron™

The RackSaver® BladeRack™, powered by AMD Opteron™ processors, is designed to support a wide variety of enterprise solutions...including yours.

Over the past 10 years, RackSaver has provided customers in every industry from Entertainment to Oil and Gas with the most powerful, reliable, and scalable high performance clusters available on the market today. Our customers understand that by increasing density in their server room they improve productivity while lowering their overall total cost of ownership. In addition, RackSaver offers complete onsite enterprise-level support 24 hours a day, 7 days a week to maximize your investment protection. Discover why industry leaders like Industrial Light and Magic, NVIDIA, Shell, WesternGeco, and Novartis rely on RackSaver for all of their high performance computing needs.



THE NEMESYS

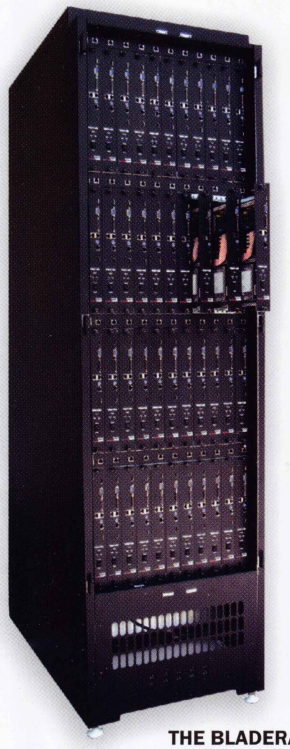
Did we mention that we sell high-end customizable graphics workstations too?

RackSaver has launched a new line of high-end graphic workstations. NemeSys™, powered by the NVIDIA Quadro® FX family, takes the leading design and digital content creation applications to a new level of interactivity. For the first time, styling and production rendering become integral functions of the design workflow, shortening the production process and enabling faster time to market.

THE HIVE CLUSTER

Need a blade solution that can fit under your desk at work and provide more power than ever before?

The HIVE™ houses up to 11 blades that can accommodate a variety of options including high-end servers, SAN, NAS-based storage, along with InfiniBand®. Both the 8 blade and 11 blade versions are fully compatible with RackSaver's RS-1100VM BladeRack line of products. The HIVE's innovative cooling technologies are specifically designed for the needs of small rack systems and also make them suitable for a wide range of environments including military, industrial or a simple office setting.



THE BLADERACK RS-1100VM

RackSaver®
Leader in High-Density Servers!

Start configuring your company's high performance supercomputing solution now by visiting RackSaver online at www.racksaver.com or call one of our Account Managers at (888) 942-3800.

Copyright ©2003, RackSaver Inc. All Rights Reserved. RackSaver, the RackSaver logo, BladeRack, and NemeSys are trademarks or registered trademarks of RackSaver Inc. Intel and Xeon and combinations thereof are trademarks of Intel Corporation or its subsidiaries in the United States or other countries.

ANGSTROM
Microsystems

AMD Opteron™ Specialists

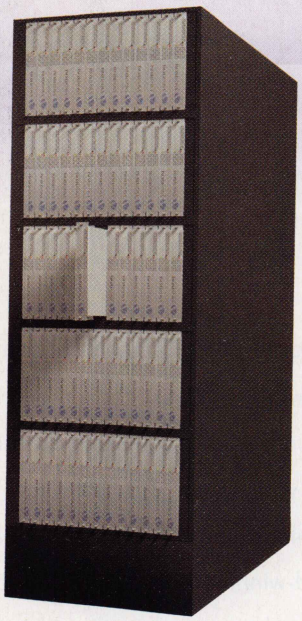
Largest AMD Opteron™ cluster for National Institute of Health installed at the University of Utah.

Over 1,000 CPUs!

AMD Opteron™

World's Densest Server Rack:

Titan64 SuperBlade™
130 SERVER BLADES - 260 AMD Opteron™ and AMD Athlon™ MP Processors in ***One Rack***!



As clusters become the standard in supercomputing solutions, more servers are required to number crunch larger and larger datasets. Titan64 SuperBlades™ solve key problems in the growth of these clusters in the following fashion:

Increased density: 128 server nodes-plus 2 management nodes per rack

Easy to maintain: Servers are hot-swappable and easy to remove

Easy to manage: System administrator can remotely manage multiple racks simultaneously

Integrate with existing datacenters: Customer specified power connectors, front-to-back cooling

Flexible design: Option to reduce density of blades and increase 19" rack space

Improve reliability: Large clusters exhibit problems not found in those that are small

Titan64 SuperBlades™ solution:
- Reduce fan count, use large fans (no small fans or blowers),
- Eliminate most cables, and provide a select indicator for blades to be serviced,
- LinuxBIOS enabled!

Titan64 SuperBlades™ more cost-effective than 1U servers

Titan64 Creator™ Workstations

High Performance Computing

The **Titan64 Creator**™ is a series of high-performance workstations with advanced 8x AGP graphics capabilities integrated with the powerful AMD Opteron™ platform to provide a content creation engine for animation, 3D modeling, and CAD.

Cutting-Edge Performance: 8x AGP graphics performance on Microsoft® and Linux® platforms running Maya®, XSI® and many other graphics intensive creation applications.

Ideal for NVIDIA® Quadro® FX 1000 by PNY®

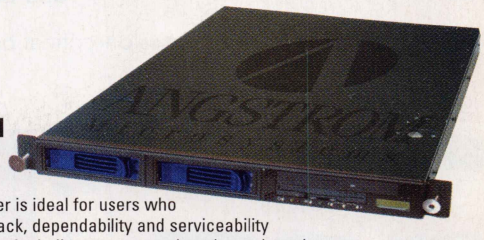
Ultra-Cool Design: Chassis design based on cooling, functionality and ease of access without sacrificing the looks

Focus On Content Creation: Our workstations and servers have created and rendered movies such as X-Men 2™, DAREDEVIL™, Scooby Doo™ The Movie and MIB II™

Proven Quality: Our proven track record for exceptional quality has enabled our clients to deploy mission-critical systems world-wide

* We offer the best price performance in the industry.

AMD Opteron™- based Enterprise Solutions



The **Titan64 Enterprise** server is ideal for users who demand high CPU density per rack, dependability and serviceability for a broad range of applications including storage, web and portal serving, application development, compute farms, and security. Designed for mission-critical 24 x 7 x 365 environments, the Titan64 Enterprise™ delivers on the promises of 64-bit technology for the enterprise today.

The LinuxBIOS Revolution – ideal for blade solutions

:: **Fast boot time (less than 5 seconds)**

:: **Improve quality and reduce cost by eliminating HDD**

:: **Remote server management**
- BIOS updates • Console access • Environmental monitoring
- Debugging failed server

:: **Easy to patch O/S and upgrade server software**

ANGSTROM
Microsystems

Angstrom Microsystems, Inc. • 27 Drydock Ave • Boston, MA 02210
• 1-877-4-ANGSTROM • www.angstrom.com • sales@angstrom.com

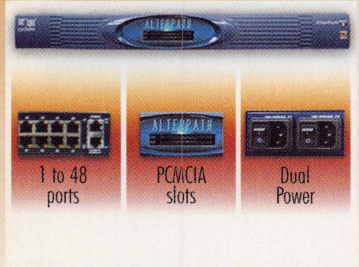
Console Management

Power Management

Command Your Network With Cyclades

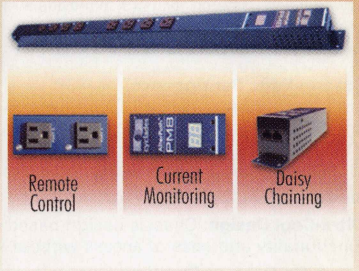
Selecting the right components for your network is often a challenging decision. With our AlterPath™ PM8, you can remotely re-boot your system with just a few mouse clicks. By integrating the AlterPath™ PM8 with our award-winning AlterPath™ ACS, you combine power and console management. Now you can command all your infastructure with secure authentication and bulletproof encryption as demanded by todays' mission-critical applications all in one single session - anytime, anywhere.

AlterPath™ ACS



Integrated Power and Console	Console and power control from one session, no need to memorize ports and addresses
Security	SSH v2, strong authentication, encryption and IP filtering on both power and console access
Daisy Chain	Daisy chain power distribution units to control any number of devices from a single serial port

AlterPath™ PM8



"Best Hardware for Linux since 1995"

www.cyclades.com/lja
1.888.CYCLADES
sales@cyclades.com

cyclades
Everywhere with Linux

©2003 Cyclades Corporation. All rights reserved. All other trademarks and product images are property of their respective owners. Product information subject to change without notice.

LINUX JOURNAL
DECEMBER 2003 ISSUE 116

COVER STORY

36 ULTIMATE LINUX BOX
Whether you're building or buying your ultimate machine, you'll be happy to know that not only has AMD rocked the PC world with a new 64-bit architecture—motherboard and Linux distribution support is keeping up with a vengeance. We like AMD64 in servers, but we love it in a workstation. The rest of the hardware market hasn't been napping, though, as we discover when we upgrade to Serial ATA RAID storage, 24-bit/96kHz audio and workstation-class 3-D graphics.

FEATURES

36 ULTIMATE LINUX BOX
We load up the hottest new 64-bit architecture with the hottest new 3-D and storage hardware.
GLENN STONE

42 DVD PLAYERS
Dave compares and tests the best DVD-playing software.
DAVE PHILLIPS

50 DVD AUTHORING
Got ideas for TV-based information design, games and Easter eggs? Make them work on DVDs you create with this versatile software.
IAN POINTER

60 MANAGING AUDIO WITH PD
Make your band sound like a symphony orchestra from Uranus with this drag-and-drop sound processing tool.
PETER TODD

68 FLOPPIES FOR THE NEW MILLENNIUM
A practical guide to setting up and working with USB key chains.
RICK MOEN

INDEPTH

84 EMBEDDING PERL IN MYSQL
Hey! There's a mistake in this article—Perl code in the middle of an SQL SELECT! Don't you people even read this stuff?
BRIAN AKER

88 CROSS-PLATFORM CD INDEX
We liked this easy JavaScript-based CD-ROM search system so much, we used it on the 1994–2002 *Linux Journal* archive CD.
SHAWN P. GARBETT

91 DVD TRANSCODING VIA LINUX METACOMPUTING
Strategies for converting MPEG-2 video from DVDs to MPEG-4 for next-generation home media applications.
F. J. GONZALEZ-CASTAÑO, R. ASOREY-CACHEDA, R. P. MARTINEZ-ALVAREZ, E. COMESAÑA-SEIJO AND J. VALES-ALONSO

EMBEDDED

30 DRIVING ME NUTS
I2C Drivers, Part I
GREG KROAH-HARTMAN

TOOLBOX

12 AT THE FORGE
Integrating E-mail
REUVEN M. LERNER

18 KERNEL KORNER
Allocating Memory in the Kernel
ROBERT LOVE

24 COOKING WITH LINUX
Put Another Nickel in...
MARCEL GAGNÉ

COLUMNS

77 LINUX FOR SUITS
Free Business
DOC SEARLS

96 EOF
Give TCPA an Owner Override
SETH DAVID SCHOEN

REVIEWS

66 INSIDE THE SECURITY MIND: MAKING THE TOUGH DECISIONS
PAUL BARRY

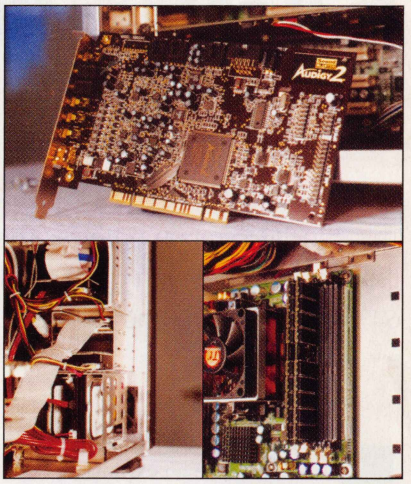
79 LINDOWS MOBILEPC/SERVELINUX ENOTE
STEVE R. HASTINGS

82 LINDOWS 4.0
STEVE R. HASTINGS

DEPARTMENTS

4 FROM THE EDITOR
6 LETTERS
8 UPFRONT
73 ON THE WEB
74 BEST OF TECHNICAL SUPPORT
76 NEW PRODUCTS
81 ADVERTISERS INDEX
95 MARKETPLACE

COVER PHOTO:
BETHANY PASEMAN



Besides AMD processors, we chose the best audio, video and storage parts for this year's Ultimate Linux Box, page 36.

NEXT MONTH

KERNEL AND SCALABILITY

In the October 2003 issue, Paul McKenney covered the basics of read-copy update, a powerful new method for handling synchronization of data structures on SMP systems. Next month, he's back with Dipankar Sarma and Maneesh Soni to explain how RCU is making a real difference in the 2.6 directory-entry cache, or dcache. Watch those scalability numbers go up, up, up.

ioctls are a way to handle miscellaneous stuff about I/O devices, like the Ethernet link light and the Eject button on the CD-ROM drive. Lisa Corsetti explains how to find the right ioctl and write a simple program to find out facts about hardware and make it do what you want.

Don't lie awake at night wondering if your hard drives are going bad, simply check your logs in the morning and see just how healthy they are. Bruce Allen explains the SMART standard for monitoring drives.

Freedom, the 64-bit Way

This special issue on entertainment puts you on the cutting edge with great media apps and a 64-bit workstation you can build.

BY DON MARTI

Freedom and DVDs, unfortunately, make an explosive combination. In the US, a judge cited the controversial Digital Millennium Copyright Act (DMCA) in a decision banning our technology journalism colleagues at *2600* from even linking to one DVD-descrambling program, DeCSS.

In other countries, though, whether you can play DVDs on your Linux box depends on whether you let your government impose a DMCA-like law. Today, the UK Campaign for Digital Rights (ukcrd.org) and other organizations are trying to prevent DMCA-like laws from taking effect in Europe.

While the legal sniping continues, we're getting more and better free, open-source DVD software for Linux. Dave Phillips covers his favorite players and some DVD-playing tips on page 42. Ian Pointer shows how to make your own DVDs with menus and background music on page 50. By the time DVD-playing software for Linux comes under attack outside the US, it won't be merely a C program on a Web site. As Linux desktop adoption continues to grow, DVD software for Linux will run on millions of desktops that users take for granted.

And, those users are voters and jurors too. A California jury acquitted Elcomsoft, a Russian software company that developed software to convert proprietary "e-books" to open formats, of DMCA charges. That's not surprising, considering that California jurors probably are used to desktop applications and don't think selecting "Save as..." from the File menu is a crime. Jury foreman Dennis Strader told the *San Jose Mercury News*, "Under the eBook formats, you have no rights at all, and the jury had trouble with that concept."

It's hard to sell a product without a demo, and freedom needs a demo too. So it's a race. Can freedom lovers deploy software that gives people a chance to experience their digital rights before governments take those rights away? Congress' attempt to impose mandatory digital rights management in the US didn't go anywhere, so the incredibly versatile effects builder Pure Data (page 60) is safe for now, as is the jukebox at *Chez Marcel* (page 24). The more people are used to doing something, the harder it would be to take away.

Gaining an ally in the struggle for freedom and justice—what better reason to build someone a Linux system? When you read Glenn Stone's article on page 36, build an Ultimate Linux Box for a family member or friend, too. The Ultimate Linux Box gets to be more of a bargain every year, even as we add 24/96 audio and move up to the 64-bit AMD64 architecture. Turn up the volume, put on some Oggs, break out the power screwdriver and enjoy this issue.

Don Marti is editor in chief of *Linux Journal*.

LINUX JOURNAL

DECEMBER 2003
ISSUE 116

EDITOR IN CHIEF Don Marti, ljeditor@ssc.com

MANAGING EDITOR Jill Franklin, jill@ssc.com

SENIOR EDITOR Doc Searls, doc@ssc.com

SENIOR EDITOR Heather Mead, heather@ssc.com

ART DIRECTOR Garrick Antikajian, garrick@ssc.com

TECHNICAL EDITOR Michael Baxter, mab@cruzio.com

SENIOR COLUMNIST Reuven Lerner, reuven@lerner.co.il

CHEF FRANÇAIS Marcel Gagné, mggagne@salmar.com

SECURITY EDITOR Mick Bauer, mick@visi.com

CONTRIBUTING EDITORS

David A. Bandel • Greg Kroah-Hartman • Ibrahim Haddad • Robert Love • Zack Brown

PROOFREADER Geri Gale

VP OF SALES AND MARKETING Carlie Fairchild, carlie@ssc.com

MARKETING MANAGER Rebecca Cassity, rebecca@ssc.com

INTERNATIONAL MARKET ANALYST James Gray, jgray@ssc.com

REGIONAL ADVERTISING SALES

NORTHERN USA: Joseph Krack, +1 866-423-7722 (toll-free)

EASTERN USA: Martin Seto, +1 905-947-8846

SOUTHERN USA: Annie Tiemann, +1 866-965-6646 (toll-free)

ADVERTISING INQUIRIES ads@ssc.com

PUBLISHER Phil Hughes, phil@ssc.com

WEBMASTER Jeff Tinsler, webmaster@ssc.com

TECHNICAL MANAGER Dan Wilder, dan@ssc.com

***LINUX JOURNAL* IS PUBLISHED BY, AND IS A REGISTERED TRADE NAME OF, SPECIALIZED SYSTEMS CONSULTANTS, INC.**

PO Box 55549, Seattle, WA 98155-0549 USA • linux@ssc.com

EDITORIAL ADVISORY BOARD

Daniel Frye, Director, IBM Linux Technology Center

Jon "maddog" Hall, President, Linux International

Lawrence Lessig, Professor of Law, Stanford University

Ransom Love, Director of Strategic Relationships, Family and Church History Department, Church of Jesus Christ of Latter-day Saints

Sam Ockman, CEO, Penguin Computing

Bruce Perens

Bdale Garbee, Linux CTO, HP

Danese Cooper, Manager, Open Source Program Office, Sun Microsystems

SUBSCRIPTIONS

E-MAIL: subs@ssc.com • URL: www.linuxjournal.com

PHONE: +1 206-297-7514 • FAX: +1 206-297-7515

TOLL-FREE: 1-888-66-LINUX • MAIL: PO Box 55549, Seattle, WA 98155-0549 USA • Please allow 4–6 weeks for processing address changes and orders • PRINTED IN USA

USPS *LINUX JOURNAL* (ISSN 1075-3583) is published monthly by SSC, Inc., 2208 NW Market Street, Suite 407, Seattle, WA 98107. Periodicals postage paid at Seattle, Washington and at additional mailing offices. Cover price is $5 US. Subscription rate is $25/year in the United States, $32 in Canada and Mexico, $62 elsewhere. POSTMASTER: Please send address changes to *Linux Journal*, PO Box 55549, Seattle, WA 98155-0549. Subscriptions start with the next issue. Back issues, if available, may be ordered from the Linux Journal Store: store.linuxjournal.com.

LINUX is a registered trademark of Linus Torvalds.

Powerful

jet
locomotive
bodybuilder
Paracel Cyclone

Reliable

dog
firefighter
parachute
Paracel Cyclone

Flexible

clay
wire
gymnast
Paracel Cyclone

Powerful Reliable *and* Flexible

Paracel Cyclone

Paracel Cyclone

For over a decade, Paracel has been providing high-performance computing solutions for government agencies, academic and research institutions, and Fortune 500 companies worldwide. These customers rely on Paracel's proven solutions and expertise for their **mission-critical applications.**

Paracel was among the first to develop systems powered by Linux clusters to solve some of the most computationally demanding and complex problems. At Paracel, we leverage our extensive knowledge and experience while working with you to **design and build a Linux cluster** specifically for your application.

We deliver complete systems pre-configured with all of the hardware and software options and tools you'll need for seamless migration, rapid deployment, and optimal performance. Whether you need eight CPUs or a thousand, for applications ranging from computational fluid dynamics to graphics rendering or life science research, Paracel can provide a **solution that meets your unique computing needs.**

Call us today for your free expert consultation at 877.321.FAST

Paracel Cyclone

The Paracel Cyclone is a turnkey Linux cluster solution, available in flexible configurations for a range of applications. Built using only high-quality components and tested rigorously, the Paracel Cyclone provides the ultimate in performance and reliability. The Paracel Cyclone is backed by a three-year warranty and Paracel's uncompromising service and support. Contact us today for your free expert consultation at 877.321.FAST

PARACEL®

Applied High-Performance Computing

www.paracelcyclone.com

CYCLONE

SCO Crimes

Following the SCO vs. IBM fiasco in *LJ* and various other publications, I have not seen any mention of possible criminal charges that could be filed against SCO and its officers should they lose their case. I would expect to see numerous counts of mail fraud to be followed up by extortion charges for sending letters demanding payment. I'm no lawyer, but I'd think racketeering charges may even be warranted.

--

Randy Bancroft

Sounds like you should be reading GROKLAW (groklaw.com), where paralegal Pamela Jones explains SCO's precarious legal situation in merciless detail.—Ed.

Correct URL for PhotoGen

Just wanted to let you know that the URL listed for my project (PhotoGen) is not correct [*LJ*, October 2003, page 10]. You guys have shawley.mpip.org, and it is supposed to be shawley.myip.org. By the way, thanks for putting my project in the mag, that rocks!

--

Jeremy

More on Audio Apps, Please

I liked the blurb about the Jahshaka video editor in the October 2003 issue. However, I'd like to see some coverage on audio multitracking software or audio software in general. Ardour is a very powerful tool for audio recording, and I think that many artists would benefit from knowing that they don't have to pay $1,000 plus for a ProTools rig to produce their art effectively. And lastly, I think *Linux Journal* is fabulous; I read every issue cover to cover. Thank you.

--

Jeremy Sherrer

Check out the Pure Data article on page 60.—Ed.

New Wi-Fi Freedom in France

I am a French *Linux Journal* reader. I found the article, "Linux Makes Wi-Fi Happen in New York City" by Doc Searls, in the September 2003 issue of *LJ*, very interesting, but I would like to comment on the following sentence: "In Europe they use 1–13, except for France, where they use 10–13." Since July 25, 2003, you can use channels 1–13 in France.

--

Romain Touze, Paris, France

Airplane Seatback System Runs Linux

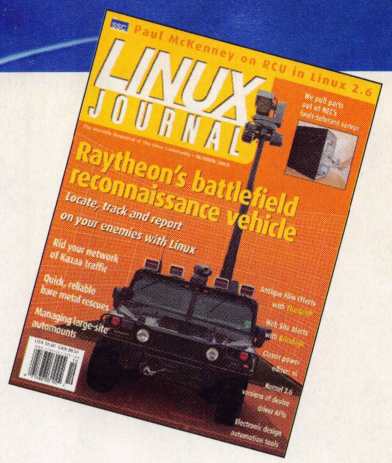
The most amazing thing happened when I was flying back from Detroit—where I had been teaching a Linux course for one of our customers—to Amsterdam late Wednesday night. The Northwest airplane was a brand-new Airbus 330, which had just been put into service the week before. It's their only Airbus so far, and it was completely new to the cabin crew. They told me that they had had their training for this aircraft based on photos from the galley.

The A330 is equipped with an in-flight entertainment system for each and every individual passenger—even in cattle class every seat has its own LCD panel, remote control, et cetera. The system is unique in the respect that it allows you to choose and start a movie at any time. So every passenger could be watching a different part of a movie at the same time. (This is actually great—sometimes you hear a giggle behind you, and five minutes later a giggle in front of you. Two people watching the same thing, only with five minutes difference in the time they started it. Happened again a few times.)

Anyway, the cabin crew did not know how to operate the system properly, so we couldn't even get our reading lights going. After they messed with it for about 30 minutes, they announced over the PA system that they would be doing a full power reset on the system. They had evidently talked to tech support via the sat phone, and that seemed to be the only resolution. And true enough, power went off and the systems started booting—Linux.

I could not believe my eyes. Here were more than 300 passengers staring at Tux's friendly smile, while Linux was booting on each and every console, in frame-buffer mode. I could see that they are using cramfs for their filesystems, but not much else before X took over. Needless to say, I was amazed. I should have brought a camera with me. That would have been a great picture for *LJ*.

Once the system was rebooted, it worked properly. Movies, multi-user games, interactive maps, surveys, you name it. You could even send e-mail (but that would cost



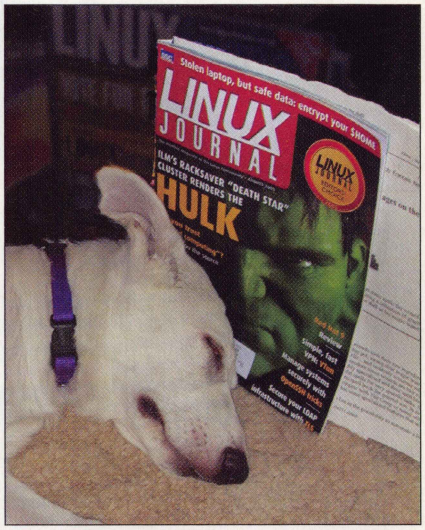
$2.50 US, and you needed to type it using a joystick-like device—not very fast). Currently, I think the range of applications is actually quite limited. The only multi-user games they've got right now are multiplayer *reversi*, and some sort of trivia quiz. But can you imagine playing *Quake* at 30,000 feet with the whole passenger load or getting a true chess competition going on in-flight? Anyway, just wanted to let you know that our favorite OS is not only used on earth and in space, but in between these two as well, and it's keeping a lot of people entertained.

--

Wouter Liefting, IBM Learning Services

Another Happy Reader

I just wanted to let you know that *Linux Journal* continues to be my favorite source of Linux-related information. And, it has become my dog Gracie's magazine of choice as well. As she's only three months old, she does get tired easily and has to take an occasional nap.

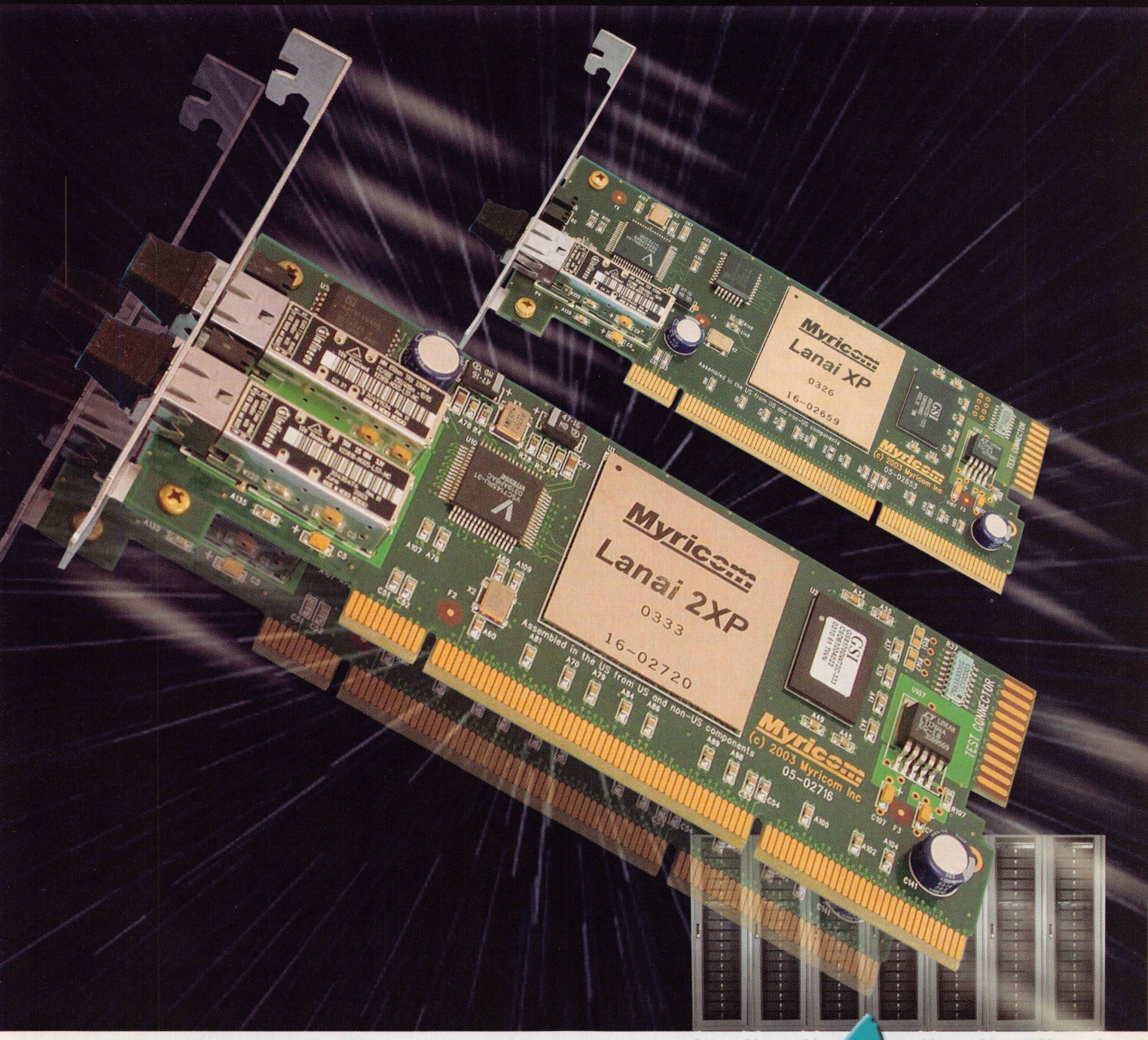


--

Larry Varney

LETTERS CONTINUED ON PAGE 81

ASPEN SYSTEMS PROVIDES THE WORLD'S BEST BEOWULF CLUSTER SOLUTIONS



Aspen Systems offers Beowulf clusters and high-performance computing (HPC) solutions with Myricom's Myrinet-2000 PCI-X cards: the low-cost M3F-PCIXD card, and the high-end M3F2-PCIXE card. Aspen Systems' customized Beowulf clusters and HPC solutions with the new E card offer 2 fiber ports per card for double the bandwidth, so now you can have up to 1GByte/s — perfectly balanced with the 1GByte/s PCI-X bus — for a double-speed and highly reliable connection.

Myricom's D and E card PCI-X interfaces were designed to work efficiently with GM-2 and MX Myrinet software support, supplied open-source by Myricom. GM-2 is the latest evolution of a message-passing system that is used in thousands of clusters worldwide. MX is a new message-passing system that is optimized for MPI, VI, and Sockets middleware. The table shows the user-level microbenchmarks for GM-2 and MX on D and E cards in dual-Xeon hosts.

For additional information about Aspen Systems Beowulf clusters and HPC solutions with the Myrinet-2000 PCI-X cards, contact Aspen Systems' team of sales engineers by calling toll-free (U.S. and Canada only) at **800.992.9242** or **303.431.4606** or via the Internet at **www.aspsys.com**

Aspen Systems Inc.

D card Microbenchmark	*Units*	*GM*	*MX*
One-way data rate	MB/s	248	248
Bidirectional data rate	MB/s	489	489
API-level latency	µs	6	4
MPI latency	µs	6.5	4
Host-CPU overhead per message	µs	<0.5	<0.5
E card Microbenchmark	***Units***	***GM***	***MX***
One-way data rate	MB/s	>480	>480
Bidirectional data rate	MB/s	>950	>950
API-level latency	µs	5	3.5
MPI latency	µs	5.5	3.5
Host-CPU overhead per message	µs	<0.5	<0.5

Myricom AUTHORIZED RESELLER

diff -u: What's New in Kernel Development

Bryan O'Sullivan has produced **Netplug**, a dæmon that monitors whether network cables are plugged in and responds by bringing up or shutting down the system's network connection. This is useful for laptops and other systems that move around a lot or that are part of a Beowulf or other cluster. Netplug performs a similar task to the existing **ifplugd** Project by **Lennart Poettering** and others, and there may be some movement toward integrating the two projects.

The **/proc/kcore** interface to system memory may be going away in 2.6 because of the difficulty in maintaining it across the range of architectures and even across multiple versions of a single architecture. **Linus Torvalds** feels that few people actually use the interface, judging from the lack of complaints whenever it breaks; he believes it may be more trouble than it's worth. It turns out, however, that some developers do use the interface for debugging purposes, although they all agree a much better solution would be to include a proper debugger in the kernel itself. Linus always has been reluctant to do this, because he feels it lowers the quality of developers' debugging efforts.

Several developers have lodged complaints with Linus Torvalds, because he occasionally modifies changelog entries for patches that already have been accepted into the **BitKeeper** repository. Some developers call this censorship, some say it could lead to legal hassles, and some say it avoids proper version control because the history of the changelog entry is lost. But Linus feels it is good practice to make sure all changelog entries are well formed and accurately describe their corresponding patches. And, it was he who asked **Larry McVoy** to add the **bk comment** command, which easily allows such modifications.

NFS behaves a bit differently in 2.6 from the way it behaved in 2.4, resulting in a true incompatibility between the two stable series. The problem seems to boil down to a `#define NFSEXP_CROSSMNT` expression that always had a slightly inaccurate meaning, one that was corrected in the 2.6-test kernels. Changing it back would mean re-introducing the inaccuracy, so it looks as though the incompatibility will stick around. The incompatibility occurs when third-party code includes the header file containing the affected #define. A simple workaround exists to fix all affected source code.

Junio C. Hamano brought **Phillip Lougher**'s **SquashFS** compressed filesystem code from 2.4 up to the 2.6-test tree. SquashFS appears to be the most promising compressed filesystem currently available for Linux, although it still provides only one-time write access; thereafter only read access is available. Full read-write access may be added in the future, however. Meanwhile, its predecessor, CramFS, is looking for another maintainer. **Daniel Quinlan** remains the official maintainer, but he would prefer to step down if a suitable replacement can be found.

—ZACK BROWN

They Said It

They are smoking crack.

—LINUS TORVALDS, REFERRING TO THE SCO GROUP'S CLAIM THAT LINUX CONTAINS A MILLION LINES OF CODE COPIED FROM UNIX.

Linux is not a product. Rather, Linux is a collection of software components, individually crafted by thousands of independent hands around the world, with each component changing and evolving on its own independent timetable.

To think of Linux as a product is to freeze an inherently dynamic thing in time and to close something that is inherently open. It cannot be done without losing something—and something significant at that.

No, Linux is not a product. It is a process.

—IAN MURDOCK, news.com.com/2010-1071_3-5057321.html)

When I was in 5th grade, an argument about personal calculators broke out in the Letters to the Editor column of my local newspaper, an argument conducted with religious fervor. On one side were those who felt that calculators were, on the whole, a good tool and that students should be taught to use them. On the other side were those who felt that calculators were a crutch, that their use would pollute the minds of impressionable youth and that their appearance was a sure sign of The End Times.

What both optimists and pessimists believed, however, deep down, was that their opinions mattered. "Someday", each of them thought, "someone is going to ask me what we should do about these here calculators." What the adults didn't understand, but me and my 5th grade posse did, was that calculators, having arrived, were never going away.

—CLAY SHIRKY, www.corante.com/many/20030701.shtml#46771

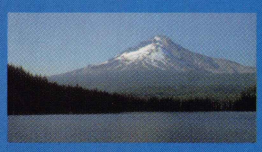
Asymptopia Flash Card: www.asymptopia.com

Want to study while you're working at your computer? This program puts flash cards up at predetermined intervals so you can practice while you're waiting for downloads or browsing the Web. Making new cards does require knowledge of LaTeX, but you can follow the example cards. Requires: libXm, libXpm, libXext, libXt, libSM, libICE, libX11, libasymptopia, libm, libstdc++, glibc, libgcc_s, libXp, libdl, libtiff, libpng12, libjpeg, libz and latex.

—DAVID A. BANDEL

The Compilers and Tools for Scientists and Engineers.

C, C++, F77, F90 and HPF • 32-bit and 64-bit Linux
Optimized for Intel and AMD CPUs • 64-bit support on AMD64
Workstation, Server and Cluster configurations • Fast compile times
Native OpenMP • Native SMP auto-parallelization
Native distributed-memory parallelization • Cache tiling
Function inlining • SSE/SSE2 Vectorization • Loop unrolling
Interprocedural optimization • Profile-feedback optimization
Large file support on 32-bit Linux • 64-bit integers and pointers
F77 pointers • Byte-swapping I/O • VAX and IBM extensions
OpenMP/MPI/threads debugging • OpenMP/MPI/threads profiling
Interoperable with g77/gcc/gdb • PDF and printed documentation
Electronic purchase, download and upgrades • Tech support
Network-floating licenses • Academic and volume discounts



The Portland Group™ Compiler Technology

www.pgroup.com ++01 (503) 682-2806 sales@pgroup.com

The registered trademarks and marks are the property of their respective owners.

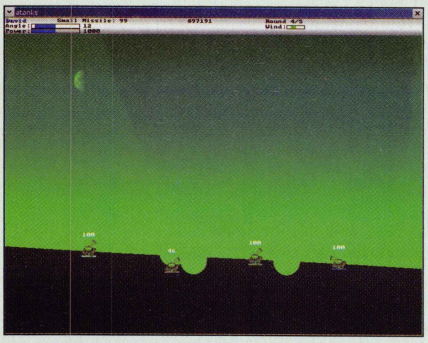
STMicroelectronics

Krystal Drop: krystaldrop.sourceforge.net

Krystal Drop is a game similar to *Tetris* in that you line up several krystals in a row. However, the alignment is vertical, and you have to pop all the similar colored crystals around the vertically aligned krystals. This is more difficult than it sounds. My games last only about a minute, but maybe I need a lot more practice. Requires: libSDL, libpthread, libSDL_image-1.2, libSDL_ttf-2.0, libGL, libSDL_mixer-1.2, libxml2, libz, libstdc++, libm, libgcc_s, glibc, libX11, libXext, libdl, libjpeg, libpng, libfreetype, libvorbisfile, libvorbis, libogg and libsmpeg.

—DAVID A. BANDEL

Atanks: atanks.sourceforge.net



Next to adventure, this game or a variation of it has to be about the oldest computer game. I remember playing a game like this years ago in which two artillery guns lobbed crude-looking bombs at each other. But even before that, in the days when I was using punched paper tape and an acoustic 75-baud modem to connect to a mainframe, I remember playing a game where you input gravity and wind, then set angle and round velocity and tried to hit a computer gun before it hit you. Although light-years beyond either of those primitive games, *Atanks* is basically the same, only with more and better weapons and shields. Requires: liballeg, libpthread, libXext, libX11, libdl, libstdc++, libm, libgcc_s and glibc.

—DAVID A. BANDEL

LJ Index—December 2003

1. Fine, in Euros, charged by the city of Munich against The SCO Group for making unsubstantiated claims about Linux: 10,000
2. Number of PC Club stores expected to stock computers with LindowsOS preinstalled: 51
3. Number of years ibiblio.org has been distributing and hosting with Linux: 12
4. Percentage of surveyed UK Web surfers who would rather give up mobile phones than the Web: 40
5. Estimated amount in billions of dollars of the Linux server market in 2002: 2
6. Estimated amount in billions of dollars of the Linux server market in 2007: 15
7. Projected Linux installation percentage on PCs sold in India by March 2004: 10
8. Indian percentage of the offshore IT services market: 60
9. Estimated annual size in billions of dollars of the offshore IT market: 16
10. Rupert Murdoch's predicted approximate price for a personal video recorder (such as a TiVo) within one year: 0
11. Price Echostar (DishTV) says it plans to charge for its PVR if customers sign up for two years, plus certain programming packages: 0
12. Average annual earnings (EBIT) percentage growth target for Rupert Murdoch's News Corp.: 20
13. Billions of dollars in worldwide B2B transactions in 1999: 145
14. Projected trillions of dollars in B2B transactions in 2004: 7.29
15. US percentage of B2B transactions in 1999: 39
16. Projected US percentage of B2B transactions in 2004: 39
17. Percentage of all developers surveyed who expect to write Linux applications in the next year: 59
18. Percentage of Linux developers who say the SCO vs. IBM lawsuit will affect their plans: 6
19. Microsoft's projected Linux server growth percentage from July 2003–June 2004: 24
20. Microsoft's projected Windows server growth percentage in the same period: 9.5

Sources

1: *Heise Online*

2: Lindows.com

3: ibiblio.org

4: *the inquirer*

5–6: Giga Information Group, via Veritas

7: Red Hat

8, 9: Gartner Group

10–12: Reuters

13–16: Gartner Group

17, 18: Evans Data Corp.

19, 20: Jupiter Research

Rackspace — Managed Hosting backed by Fanatical Support.

Servers, data center and bandwidth are not the key to hosting enterprise class Web sites and Web applications. At Rackspace, we believe hosting is a service, not just technology.

Fanatical Support™ is our philosophy, our credo. It reflects our desire to bring responsiveness and value to everything we do for our customers. You will experience Fanatical Support from the moment we answer the phone and you begin to interact with our employees.

Fanatical Support has made Rackspace the fastest growing hosting company in the world. Call today to experience the difference with Fanatical Support at Rackspace.

rackspace™
MANAGED HOSTING

1.888.571.8976 or visit us at **www.rackspace.com**

Integrating E-mail

Keep users coming back to your Web site with e-mail reminders about news or discussions that interest them. BY REUVEN M. LERNER

I have been using my computer to communicate with other people for more than 20 years. What began as occasional participation in forums on local bulletin boards has become an inseparable part of my personal and professional lives. The daily flood of e-mail I receive from friends, relatives and colleagues often seems overwhelming—until I spend time without access to e-mail, at which point I realize exactly how important and useful it is.

This month, as I continue to unpack at my new home in Chicago and fight the bugs and glitches that make it difficult for me to continue discussing Bricolage, I look at a number of different issues having to do with e-mail in the modern era—Web/mail integration, mail/database integration and even fighting spam at the SMTP level.

Integrating the Web and E-mail

During the last two years, at least a dozen clients have asked me to set up Web-based forums. Indeed, it is rare to find a large, modern Web site that does not include a section for user feedback or participation. The question is, how do you incorporate forums into your site? In many cases, the answer depends on the type of site you're running.

If you're using a large toolkit, such as OpenACS, Zope or PHPNuke, at least one forum package is available for easy installation into your site. Not only is the look and feel of it closely integrated into the static pages and other applications, but these toolkits allow the same users and permissions you have set on the rest of the site to have access to the forums. In other words, you don't have to make someone a site administrator and a forum administrator separately; anyone with administrative privileges on the site is able to run the forums or any other application without further configuration.

Alternatively, you can install a separate Web forums package using the server-side technologies available on your server. For example, if PHP is available, you can install the Phorum package. It won't be integrated completely into the rest of your site, but Phorum is a powerful and stable package and works with both MySQL and PostgreSQL. There are many options from which to choose, with underlying technologies ranging from PHP and JSP to plain-old CGI programs written in Perl.

You also can go the proprietary route and license a package such as WebCrossing. I've used WebCrossing for a few clients, and although this package offers many more features than do open-source products such as Phorum, the differences are not great enough in most cases to warrant spending the money, let alone learning a package that cannot be modified or improved.

Finally, you can roll your own forums package, as I did in this column several years ago. Writing a set of programs that implement Web-based forums isn't difficult, but the time and effort in developing and debugging would be spent better learning an existing package. But for anyone who has a bit of experience with the Web and databases, producing a Web forums package means creating only a few tables (for users, messages and threads) and then providing people with the ability to insert new records (postings) into each thread.

All of these systems are more than sufficient for a typical small- to medium-sized Web site. And even if your site grows large, with thousands of messages posted by hundreds of users, it will unlikely tax any of these systems. That's because these systems all use relational databases to store messages, and even the smallest and simplest modern database server can handle thousands of transactions per day.

In an increasing number of cases, however, a simple Web-based forum is not enough. Although people might be willing to look through Web-based forums, they probably are not going to return to them day after day to keep track of the discussion. Whereas e-mail is a push medium in which the information is sent to you, forums are a pull medium, where new messages wait for you to request them.

Push and pull is not a new issue; those of us who remember the pre-Web Internet know that discussions used to be divided between mailing lists and Usenet newsgroups. The solution was to create various mail-to-Usenet gateways, many of which continue to be used today. Indeed, you can keep track of the latest bugs in GNU Emacs by subscribing to the bug-gnu-emacs mailing list or by reading the gnu.emacs.bug newsgroup on Usenet. The two are equivalent, facilitated by gateway software that transmits messages from one system to the other without unnecessary duplicates.

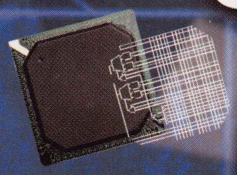
Do any such systems exist for Web-based forums? The answer is a tentative yes. After all, it's easy for a Web/database system to send mail to a list of e-mail addresses when a new message is posted. It is not much more difficult to create a sophisticated system of notifications, as provided in the OpenACS forums package, that allows users to subscribe to a particular forum or thread within a forum—and then choose to receive updates immediately or on a daily or monthly basis. In other words, the forum software is capable of creating an e-mail digest, in the same manner that list software can.

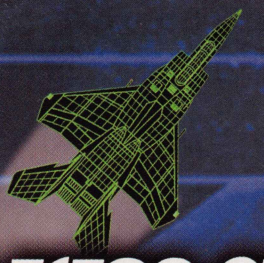
Posting Mail

Things get a bit trickier, however, when you want to let forum participants use their own e-mail software to post new messages to the list. In particular, several issues must be addressed:

- Security and permissions—should anyone be allowed to post or reply to the forum, or only members? If access to the forum is somehow restricted, it is necessary to keep track of the allowed e-mail addresses. Of course, the fact that it's so easy to forge e-mail headers means it's impossible to verify completely that a posting truly is coming from a subscriber

High-Performance Blade Servers

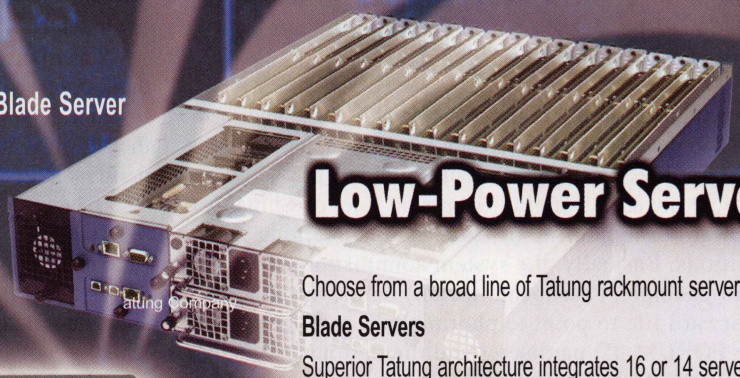




www servers

e commerce

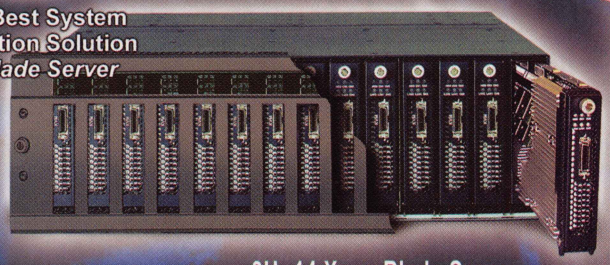
2U- 16 PIII Blade Server



Low-Power Servers

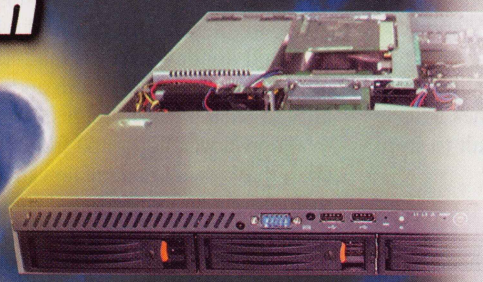
PRODUCT EXCELLENCE AWARDS
LINUXWORLD CONFERENCE & EXPO

2003 Best System Integration Solution
3U Blade Server



3U- 14 Xeon Blade Server

Point to tsti.com



1U Dual Xeon Rackmount

Choose from a broad line of Tatung rackmount servers.

Blade Servers

Superior Tatung architecture integrates 16 or 14 server blades in a 2U or 3U chassis. That form factor can deliver 336 PIII or 196 Xeon servers in a single 42U height rack while using significantly less power than comparable systems and eliminating more than 80% of the wires.

Each hot-swappable blade is powered by the latest low-power Intel® ULV PIII or LV Xeon processor and has built-in dual gigabit ethernet switches and advanced management software.

1U Xeon Servers

High Performance single or dual Intel Xeon processor rackmount servers are now shipping. Configure the 1U with up to four hard drives and a 2U or 3U with six hard drives or more.

Tatung is building rackmount servers now. Call today and have us build a system for you.

The Intelligent Choice in Server Solutions

TATUNG®
SCIENCE & TECHNOLOGY, INC.

USA: 1-800-659-5902 • www.tsti.com
Europe: www.tatung.nl
Asia: www.tatungserver.com

Intel is a registered tradmark of Intel Corp.
Tatung is a registered trademark of T

rather than from a worm or virus posing as a subscriber.

- MIME—most popular e-mail programs, especially Microsoft Outlook, send mail by default with one or more attachments. Forum software must be intelligent enough to handle postings sent in this format, stripping out the HTML and attachments.

- Size—if the mail-to-forum gateway is not intelligent enough to weed out extremely large postings, someone could mount a denial-of-service attack against your Web site by submitting extremely large postings. The software needs to be smart enough to filter this mail out, allowing the site administrator to limit the size of user postings.

- Threading—most forum software keeps related postings together, either by subject title or by keeping track of which posting was a reply to another posting. It's difficult but somewhat possible to keep track of this when combining the Web and e-mail, because each medium uses a different mechanism for keeping track of such things.

I have seen a number of different ways to handle these and other issues. To date, I haven't seen any to my complete liking.

Phorum provides a script called phorummail that is designed to be invoked from the command line, presumably from a .forward or .qmail file or from a file containing mail alias definitions. Basically, the system administrator creates an alias (for example, apartments-forum) on the system and sets the .forward file to point to phorummail, passing the mandatory FORUM_ID parameter and the optional PATH_TO_FORUM parameter. Once you have done that, anyone sending mail to apartments-forum@*your.site* can post to the forum. Obviously, permissions have to be set appropriately in order for phorummail to work.

The problem is there isn't much security on the posting; although, if the forum is moderated, postings that originated by mail are marked as unapproved until a moderator reviews them. Threading is taken care of in a number of clever ways, but there doesn't seem to be any provision for handling MIME or mail-bombing attacks. In other words, Phorum handles mail-to-forums as you might expect but without fancy features.

The OpenACS forum software includes a more sophisticated system based on qmail, in which every outgoing notification message is sent from a unique identifier. This means a reply to an e-mailed message is submitted to the forum in question. The fact that OpenACS uses e-mail addresses as login names adds a tiny bit of security, but the fact remains that it's difficult to stop people from forging messages.

Mailing Lists and Databases

The best solution I can offer is to turn things on their head by making a Web-based forum an offshoot of an existing e-mail list. ezmlm, the mailing list package written by qmail author Dan Bernstein, has a set of extensions known as ezmlm-idx that, among other things, allows the subscriber list to be stored in either MySQL or PostgreSQL. When configuring the list, the administrator also configures some database tables and then points ezmlm to those tables.

This means that a good Web developer can create an e-mail list and then mirror that list to the Web. Anything coming from the Web appears to come from the currently logged-in user, who presumably had to log in to the Web-based forum application. Anything coming from an actual user goes through the same checks that ezmlm normally applies.

Mailman, a mailing list program that works with all MTAs (including qmail, Sendmail, Postfix and Exim) and that is undergoing continuous and impressive development, doesn't yet seem to have any hooks for storing its users in a relational database. It does store them in relatively safe, easy-to-use Berkeley DB files, though, which make it easy for a Web-based forum package to read them from there.

Several problems arise from handling forums as if they were e-mail lists, beginning with the issues of threading, which, as mentioned above, are different in e-mail and the Web. Add to this the fact that many forums allow users to add highlighting and attachments and even edit their own postings, and it quickly becomes obvious that the marriage is going to be difficult—an impedance mismatch, if you will, between the two media.

Nevertheless, the functionality is useful enough for a large number of people who are willing to ignore the fringe issues. Instead, they focus on increasing the number of participants in their forums and in giving users a choice regarding how they participate in these forums.

qpsmtpd

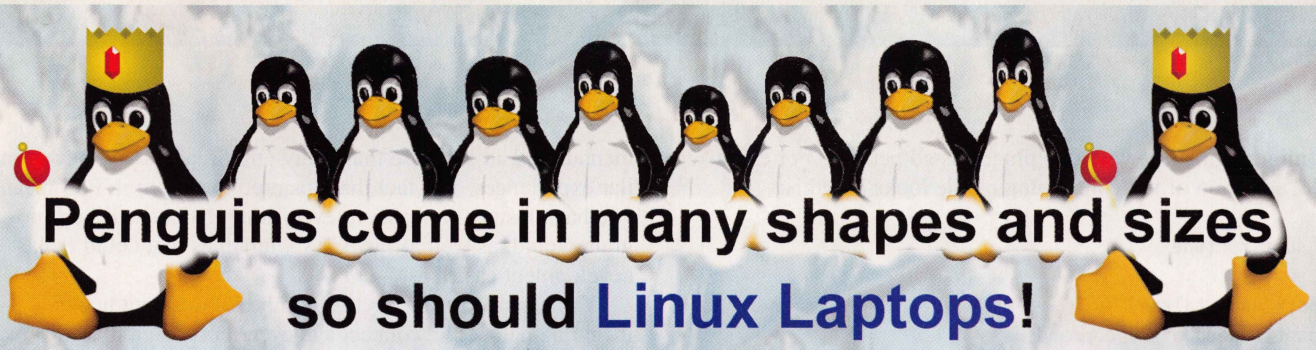
Like many of you, I am plagued by a torrent of spam on a daily basis. SpamAssassin, the open-source tool for analyzing and categorizing incoming e-mail, has proven to be an excellent ally in my fight against spam. And if you ever have run an e-mail list, you undoubtedly have discovered that e-mail worms don't discriminate; they post to lists as easily as they send mail to individuals.

Although system administration and SMTP servers are a bit off-topic for this column, I feel compelled to sing the praises of qpsmtpd, the open-source SMTP server developed by Ask Bjoern Hansen. qpsmtpd originally was designed for use with qmail, but now it apparently is able to work with other MTAs, including Sendmail and Postfix.

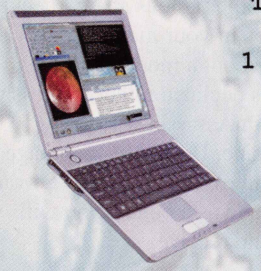
Why would you want to insert qpsmtpd instead of the default qmail-smtpd? If you're a Perl hacker, the reason to switch is qpsmtpd is written in Perl. But if you're less of a language bigot, you still can find a lot to love. That's because qpsmtpd divides SMTP's mail-sending routines into a number of stages and allows you to add your own hooks and functionality to each of these stages.

I downloaded qpsmtpd from its home at www.develooper.com (yes, that's two O characters in a row), followed the installation instructions and was up and running in about 20 minutes. Remember that qpsmtpd is a full-fledged SMTP server, meaning it refuses to run if you have another SMTP server listening on port 25. If you have been using daemontools to ensure the SMTP server starts at boot time and stays up following that, you should double-check that no link exists from /service to the old SMTP server. Otherwise, you might end up with two competing SMTP dæmons when your machine is next restarted.

The key to qpsmtpd is its plugins, which live in the plugins subdirectory. You can add or remove plugins by modifying the

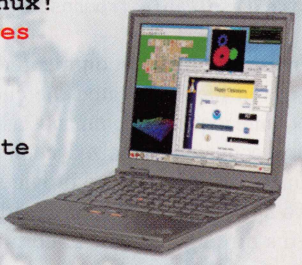


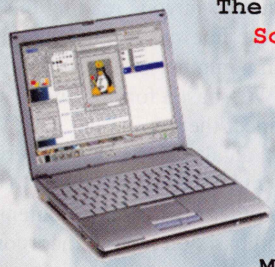
Penguins come in many shapes and sizes so should Linux Laptops!



The Meteor: 2 lb Linux!
Sharp Actius MM10
1.0 GHz Transmeta Crusoe
15 GB HD
256 MB RAM (Fixed)
802.11b Wireless
10/100 Ethernet
USB Cradle
0.5" Thin!

The Raven: 3.5 lb Linux!
IBM ThinkPad X Series
1.3-1.4 GHz P-M
20-40 GB HD
256-2048 MB RAM
APM Suspend/Hibernate
802.11b Wireless
10/100 Ethernet
FireWire
1" Thin!





The BlackPerl: 4 lb Linux!
Sony Vaio V505 Series
2.0-2.4 GHz P4-M
40-60 GB HD
256-2048 MB RAM
CDRW/DVD
FireWire
802.11b Wireless
10/100 Ethernet
Memory Stick/Jogdial

The Toucan: 5 lb Linux!
IBM ThinkPad T Series
X@1400x1050
1.6-2.4 GHz P4-M/
1.3-1.6 GHz P-M
20-80 GB HD
256-2048 MB RAM
CDRW/DVD
802.11b Wireless
APM Suspend/Hibernate

The Gazelle: 8 lb Linux!
Sony VAIO GRT Series
X@1600x1200
2.0-2.5 GHz P4-M
40-80 GB HD
256-2048 MB RAM
CDRW/DVD or DVD+/-RW
10/100 Ethernet
Memory Stick/Jogdial
FireWire

The Rhino: 7 lb Linux!
Dell Inspiron 8500/
Latitude D800
X@1920x1200
2.0-2.6 GHz P4-M/
1.6-1.7 GHz P-M
30-80 GB HD
256-2048 MB RAM
CDRW/DVD, DVD, CDRW,
or DVD+/-RW

Since 1999, EmperorLinux has provided pre-installed Linux laptop solutions to universities, corporations, and individual Linux enthusiasts. We specialize in the installation and configuration of the Linux operating system on a wide range of the finest laptop and notebook computers made by IBM, Sony, and Dell. We offer a range of the latest Linux distributions, as well as Windows dual boot options. We customize each Linux distribution to the particular machine it will run upon and provide support for: ethernet, modem, wireless, PCMCIA, USB, FireWire, X-Server, CD/DVD/CDRW, sound, APM, and more. All our systems come with one year of Linux technical support by both phone and e-mail, and full manufacturers' warranties apply. Visit www.EmperorLinux.com or call 1-888-651-6686 for details.

Successful Linux laptop solutions found for:

Sandia Nat Lab DOE NASA JHU Applied Physics Lab Max Planck Inst Woods Hole NRAO Naval Research Lab Smithsonian Institute Los Alamos Nat Lab UC Lawrence Berkeley Labs NOAA Lawrence Livermore Nat Lab US Army Research Lab U Penn U Utah U Maryland Yale Cornell Columbia U Oregon NC State Duke U Massachusetts George Mason Colorado State Caltech U Illinois Georgia Tech U Minnesota U Chicago Pace U UCSF UCSD Rutgers Princeton NYU Purdue Boston U MIT U Arizona George Washington U U Hawaii U Maine

www.EmperorLinux.com 1-888-651-6686

Model prices, specifications, and availability may vary. All trademarks are the property of their respective holders.

config/plugins file, with one plugin listed per line. For example, a portion of my config/plugins file looks like this:

```
# quit_fortune

check_earlytalker
count_unrecognized_commands 4

require_resolvable_fromhost
```

In other words, I commented out the quit_fortune plugin but have activated the check_earlytalker, count_unrecognized_commands and require_resolvable_fromhost plugins. count_unrecognized_commands takes a single numeric argument, which we have provided here.

To see these plugins or to add your own, go into the plugins directory itself. Each plugin consists of a register subroutine that attaches the plugin to one of qpsmtpd's various hooks and another subroutine that is invoked by the hook. For example, the require_resolvable_fromhost plugin begins with the following:

```
use Net::DNS qw(mx);

sub register {
  my ($self, $qp) = @_;
  $self->register_hook("mail", "mail_handler");
}
```

In other words, the register subroutine tells qpsmtpd that whenever the SMTP client invokes the mail command, the mail_handler subroutine also should be invoked. That subroutine does the following:

```
sub mail_handler {
  my ($self, $transaction, $sender) = @_;

  $sender->format ne "<>"
    and $self->qp->config
       ("require_resolvable_fromhost")
    and !check_dns($sender->host)
    and return (DENYSOFT,
       ($sender->host
        ? "Could not resolve ". $sender->host
        : "FQDN required in the envelope sender"));
  return DECLINED;
}
```

If you have done any Web development in mod_perl, this should look somewhat familiar to you. mail_handler can return DECLINED, which indicates that everything is fine and the mail can go through. Or, it can return DENYSOFT, which allows the sender to try again later. This happens because we don't want to start rejecting mail when a DNS server goes down; we're interested in punishing only spammers and others who shouldn't be sending mail directly. We also can return DENY, which rejects the mail outright.

I was able to write a new, working plugin within a few hours of downloading qpsmtpd, despite the lack of good documentation, and I'm sure that many other readers will have similar experiences. The fact that qpsmtpd is written in Perl means you have fast, easy access to everything that a usual Perl program would, as well as any CPAN modules that could make development easier.

You can attach plugins to a number of different hooks, including helo, ehlo, connect and even rcpt—each of which can perform tests of various sorts. There's even a SpamAssassin plugin for qpsmtpd, which invokes the famous spam-checking software before the message arrives in your mailbox.

I have been using qpsmtpd for about a month, and the amount of spam in my mailbox has declined rather impressively, even from the low amount that SpamAssassin was letting through. If you run your own machine, I strongly encourage you to look at qpsmtpd. It is an excellent example of how to write software to take arbitrary plugins, and as a bonus, you will receive only the mail that you should receive.

Conclusion

E-mail is a vital part of the Internet, as anyone reading this column undoubtedly knows. But as the Internet continues to expand, e-mail is being pushed in a number of different directions. This month, we looked at Web-based forums and at some of the ways in which we can connect them to e-mail lists. We also took a brief look at qpsmtpd, an alternative SMTP server designed to be highly extensible and configurable. Next month, I hope to return to Bricolage and other open-source content management systems, with an emphasis on how to incorporate a CMS into your existing Web sites.

Reuven M. Lerner, a longtime consultant in Web/database programming, is now a graduate student in Learning Sciences at Northwestern University in Chicago. As he writes this in September, he still is hoping that winters in Chicago won't be much different from what they were in Israel.

RESOURCES

You can learn more about the various open-source packages discussed this month at their respective Web sites: **www.OpenACS.org**, **www.Zope.org** and **www.Phorum.org**. The open-source databases mentioned this month are available at **www.mysql.com** and **www.postgresql.org**.

For information about qmail and ezmlm, take a look at **www.qmail.org** and **www.ezmlm.org**. The latter site includes information about ezmlm-idx, the extensions that allow you to connect ezmlm to a relational database, among other things.

qpsmtpd is available from **www.develooper.com**. A mailing list for qpsmtpd questions and hacks is available at **nntp.x.perl.org/group/perl.qpsmtpd**.

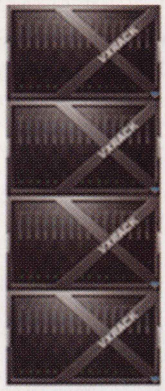
Finally, you can download SpamAssassin from **www.spamassassin.org**.

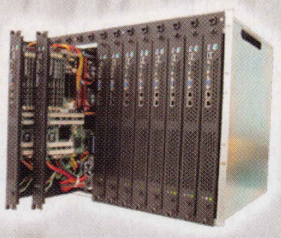
VXRACK

THE FUTURE OF CLUSTER TECHNOLOGY

NEED TO **INCREASE PERFORMANCE WITHOUT** INCREASING **COST ?**

Get VXRACK™ Cluster Technology based on the **Intel® Xeon™ Processor** and **get the price**/performance leadership **you need**.





ONE RACK
1.5TFLPS
128 VXBLADE
1.5TB MEMORY
256 INTEL® XEON™

HOW BIG IS YOURS ???

*VX*RACK™
- 8 VXCubes
- (up to 4 high and back-to-back)
- 128 VXBlade (VXB-7501)
- 256 Intel® Xeon™ Processors
- More than 32 CPU per Sft
- 51 TB of Storage
- 1.5 TB of Memory
- Patended Rack cooling system
- Environmental Thermal Sensor
- Dual Front End PDU (Optional)

*VX*CUBE™
- Rack mounted Metal Chassis
- 16 VXBlades (VXB-7501)
- 32 Intel® Xeon™ Processors
- 16 or 32 Ports Patch Panel
- (RJ45 or Infiniband Connectors
- RS-232 In-band Connector)
- Environmental Thermal Sensor
- Integrated PDU (20Amp-208V)
- Internal Gigabit Switch (Opt.)
- Internal Topspin Switch (Opt.)

*VX*BLADE™ (VXB-7501)
- VXBlade (VXB-7501)
- Dual Intel® Xeon™ Processors
- (Up to 3.06GHz)
- Intel® E7501 Chipset
- 8 or 12GB of ECC DDR memory
- Dual Extreme Active Heatsink
- Dual Intel® Gigabit Ports
- 0, 1 or 2 IDE ATA 100 HDD
- One PCI-X Slot Available
- IPMI 1.5 Management (Opt.)

*VX*BLADE™ (VXB-8870)
- VXBlade (VXB-8870)
- Dual Intel® Itanium2™ Processors
- (1GHz, 1.5MB L3 Cache)
- Intel® E8870 Chipset
- 16GB of ECC DDR memory
- Dual Extreme Active Heatsink
- Dual Intel® Gigabit Ports
- 0 or 1 IDE ATA/100 HDD
- One PCI-X Slot Available
- IPMI 1.5 Management (Opt.)

AVAILABLE NOW...CALL US FOR MORE INFOS. **COMING 12/03**

AFFORDABILITY
VXRack Extreme Density coupled with Intel Server Building Block Components offer an incredible cost-benefit compared with regular 1U Cluster solution or proprietary RISC based Supercomputers

SCALABILITY
VXRack is inherently scalable. Increased processing loads can be quickly and easily accommodated either by adding identical VXBlade & VXCube, or by scaling up the nodes with faster processors or memory

EASY UPGRADE
VXRack is easy to upgrade as workload or budget allows. Each nodes can be upgraded without affecting the other. In addition, large clusters can be conveniently built in stages on pay-as-you-go basis

RAW PERFORMANCE
VXRack can be easily built to provide virtually any level of capacity and performance. A single VXRack with 256 Xeon Processors can deliver more than 1.5Tflp (Trillion floating point operations per sec.)

FLEXIBILITY
Portions of VXRack can be assigned to different applications, allowing several separate jobs to run efficiently in parallel, or consolidated to complete a single large execution as fast as possible

UNSURPASSED QUALITY
VXRack takes advantage of Intel's 20 years of experience designing and engineering industry leading server components. Intel spends 10,000+ hours testing and validating every piece of the server stack

intel inside XEON

intel inside ITANIUM 2

CALL TODAY 866-7VX-RACK OR WWW.VXRACK.COM

CIARA TECHNOLOGIES

Copyright © 2003, Hypertechnologies Ciara Inc. All Rights reserved. Hypertechnologies Ciara Inc. is not responsible for any typographical errors. Pricing, availability or pictures are subject to change without notice. Ciara Technologies, Ciara Technologies Logo, VXRACK, VXCUBE and VXBLADE are trademarks or registered trademarks of Hypertechnologies Ciara Inc.. Intel, the Intel Inside logo, Intel Xeon and Intel Itanium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Allocating Memory in the Kernel

In this article, Robert offers a refresher on kernel memory allocation and how it has changed for the 2.6 kernel. BY ROBERT LOVE

Unfortunately for kernel developers, allocating memory in the kernel is not as simple as allocating memory in user space. A number of factors contribute to the complication, among them:

- The kernel is limited to about 1GB of virtual and physical memory.
- The kernel's memory is not pageable.
- The kernel usually wants physically contiguous memory.
- Often, the kernel must allocate the memory without sleeping.
- Mistakes in the kernel have a much higher price than they do elsewhere.

Although easy access to an abundance of memory certainly is not a luxury to the kernel, a little understanding of the issues can go a long way toward making the process relatively painless.

A General-Purpose Allocator

The general interface for allocating memory inside of the kernel is kmalloc():

```
#include <linux/slab.h>

void * kmalloc(size_t size, int flags);
```

It should look familiar—it is pretty much the same as user space's malloc(), after all—except that it takes a second argument, `flags`. Let's ignore flags for a second and see what we recognize. First off, `size` is the same here as in malloc()'s—it specifies the size in bytes of the allocation. Upon successful return, kmalloc() returns a pointer to *size* bytes of memory. The alignment of the allocated memory is suitable for storage of and access to any type of object. As with malloc(), kmalloc() can fail, and you must check its return value against NULL. Let's look at an example:

```
struct falcon *p;

p = kmalloc(sizeof (struct falcon), GFP_KERNEL);
if (!p)
  /* the allocation failed - handle appropriately */
```

Flags

The flags field controls the behavior of memory allocation. We can divide flags into three groups: action modifiers, zone modifiers and types. Action modifiers tell the kernel how to allocate memory. They specify, for example, whether the kernel can sleep (that is, whether the call to kmalloc() can block) in order to satisfy the allocation. Zone modifiers, on the other hand, tell the kernel from where the request should be satisfied. For example, some requests may need to be satisfied from memory that hardware can access through direct memory access (DMA). Finally, type flags specify a type of allocation. They group together relevant action and zone modifiers into a single mnemonic. In general, instead of specifying multiple action and zone modifiers, you specify a single type flag.

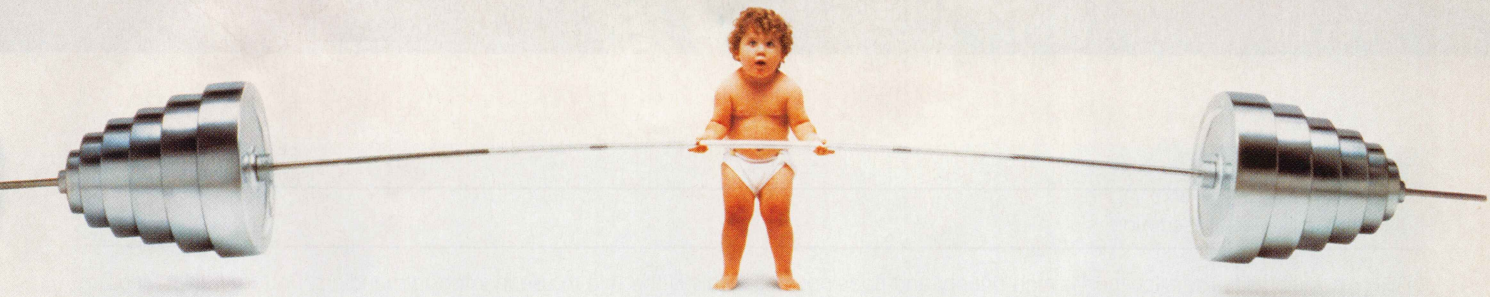
Table 1 is a listing of the action modifiers, and Table 2 is a listing of the zone modifiers. Many different flags can be used; allocating memory in the kernel is nontrivial. It is possible to control many aspects of memory allocation in the kernel. Your code should use the type flags and not the individual action and zone modifiers. The two most common flags are GFP_ATOMIC and GFP_KERNEL. Nearly all of your kernel memory allocations should specify one of these two flags.

Table 1. Action Modifiers

Flag	Description
__GFP_COLD	The kernel should use cache cold pages.
__GFP_FS	The kernel can start filesystem I/O.
__GFP_HIGH	The kernel can access emergency pools.
__GFP_IO	The kernel can start disk I/O.
__GFP_NOFAIL	The kernel can repeat the allocation.
__GFP_NORETRY	The kernel does not retry if the allocation fails.
__GFP_NOWARN	The kernel does not print failure warnings.
__GFP_REPEAT	The kernel repeats the allocation if it fails.
__GFP_WAIT	The kernel can sleep.

Table 2. Zone Modifiers

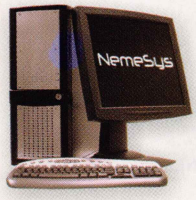
Flag	Description
__GFP_DMA	Allocate only DMA-capable memory.
No flag	Allocate from wherever available.



Experience the most computing power in the smallest footprint possible...

The RackSaver® BladeRack™, powered by Intel® Xeon® processors, is designed to support a wide variety of enterprise solutions...including yours.

Over the past 10 years, RackSaver has provided customers in every industry from Entertainment to Oil and Gas with the most powerful, reliable, and scalable high performance clusters available on the market today. Our customers understand that by increasing density in their server room they improve productivity while lowering their overall total cost of ownership. In addition, RackSaver offers complete onsite enterprise-level support 24 hours a day, 7 days a week to maximize your investment protection. Discover why industry leaders like Industrial Light and Magic, NVIDIA, Shell, WesternGeco, and Novartis rely on RackSaver for all of their high performance computing needs.



THE NEMESYS

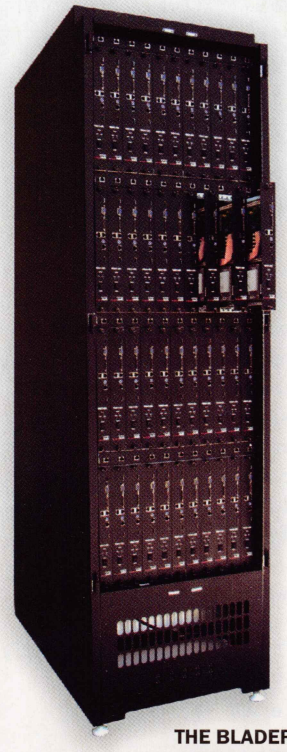
Did we mention that we sell high-end customizable graphics workstations too?

RackSaver has launched a new line of high-end graphic workstations. NemeSys™, powered by the NVIDIA Quadro® FX family, takes the leading design and digital content creation applications to a new level of interactivity. For the first time, styling and production rendering become integral functions of the design workflow, shortening the production process and enabling faster time to market.

THE HIVE CLUSTER

Need a blade solution that can fit under your desk at work and provide more power than ever before?

The HIVE™ houses up to 11 blades that can accommodate a variety of options including high-end servers, SAN, NAS-based storage, along with InfiniBand®. Both the 8 blade and 11 blade versions are fully compatible with RackSaver's RS-1100VM BladeRack line of products. The HIVE's innovative cooling technologies are specifically designed for the needs of small rack systems and also make them suitable for a wide range of environments including military, industrial or a simple office setting.



THE BLADERACK RS-1100VM

RackSaver®
Leader in High-Density Servers!

Start configuring your company's high performance supercomputing solution now by visiting RackSaver online at www.racksaver.com or call one of our Account Managers at (888) 942-3800.

Copyright ©2003, RackSaver Inc. All Rights Reserved. RackSaver, the RackSaver logo, BladeRack, and NemeSys are trademarks or registered trademarks of RackSaver Inc. Intel and Xeon and combinations thereof are trademarks of Intel Corporation or its subsidiaries in the United States or other countries.

Table 3. Types

Flag	Description
GFP_ATOMIC	The allocation is high-priority and does not sleep. This is the flag to use in interrupt handlers, bottom halves and other situations where you cannot sleep.
GFP_DMA	This is an allocation of DMA-capable memory. Device drivers that need DMA-capable memory use this flag.
GFP_KERNEL	This is a normal allocation and might block. This is the flag to use in process context code when it is safe to sleep.
GFP_NOFS	This allocation might block and might initiate disk I/O, but it does not initiate a filesystem operation. This is the flag to use in filesystem code when you cannot start another filesystem operation.
GFP_NOIO	This allocation might block, but it does not initiate block I/O. This is the flag to use in block layer code when you cannot start more block I/O.
GFP_USER	This is a normal allocation and might block. This flag is used to allocate memory for user-space processes.

Table 4. Composition of the Type Flags

Flag	Value
GFP_ATOMIC	__GFP_HIGH
GFP_NOIO	__GFP_WAIT
GFP_NOFS	(__GFP_WAIT | __GFP_IO)
GFP_KERNEL	(__GFP_WAIT | __GFP_IO | __GFP_FS)
GFP_USER	(__GFP_WAIT | __GFP_IO | __GFP_FS)
GFP_DMA	__GFP_DMA

The GFP_ATOMIC flag instructs the memory allocator never to block. Use this flag in situations where it cannot sleep—where it must remain atomic—such as interrupt handlers, bottom halves and process context code that is holding a lock. Because the kernel cannot block the allocation and try to free up sufficient memory to satisfy the request, an allocation specifying GFP_ATOMIC has a lesser chance of succeeding than one that does not. Nonetheless, if your current context is incapable of sleeping, it is your only choice. Using GFP_ATOMIC is simple:

```
struct wolf *p;

p = kmalloc(sizeof (struct wolf), GFP_ATOMIC);
if (!p)
    /* error */
```

Conversely, the GFP_KERNEL flag specifies a normal kernel allocation. Use this flag in code executing in process context without any locks. A call to kmalloc() with this flag can sleep; thus, you must use this flag only when it is safe to do so. The kernel utilizes the ability to sleep in order to free memory, if needed. Therefore, allocations that specify this flag have a greater chance of succeeding. If insufficient memory is available, for example, the kernel can block the requesting code and swap some inactive pages to disk, shrink the in-memory caches, write out buffers and so on.

Sometimes, as when writing an ISA device driver, you need to ensure that the memory allocated is capable of undergoing DMA. For ISA devices, this is memory in the first 16MB of physical memory. To ensure that the kernel allocates from this specific memory, use the GFP_DMA flag. Generally, you would use this flag in conjunction with either GFP_ATOMIC or GFP_KERNEL; you can combine flags with a binary OR operation. For example, to instruct the kernel to allocate DMA-capable memory and to sleep if needed, do:

```
char *buf;

/* we want DMA-capable memory,
 * and we can sleep if needed */
buf = kmalloc(BUF_LEN, GFP_DMA | GFP_KERNEL);
if (!buf)
    /* error */
```

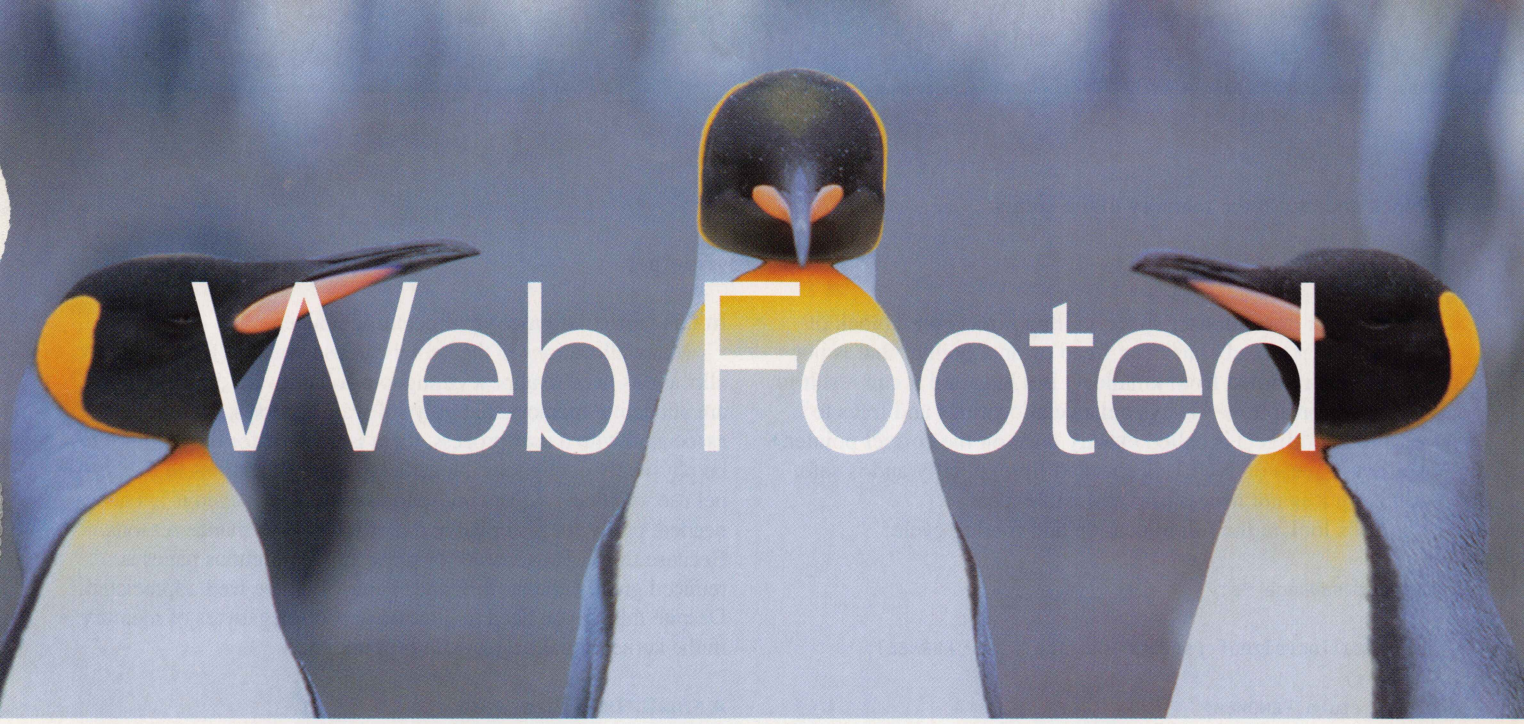
Table 3 is a listing of the type flags, and Table 4 shows to which type flag each action and zone modifier equates. The header <linux/gfp.h> defines all of the flags.

Returning Memory

When you are finished accessing the memory allocated via kmalloc(), you must return it to the kernel. This job is done using kfree(), which is the counterpart to user space's free() library call. The prototype for kfree() is:

```
#include <linux/slab.h>

void kfree(const void *objp);
```

kfree()'s usage is identical to the user-space variant. Assume p is a pointer to a block of memory obtained via kmalloc(). The following command, then, would free that



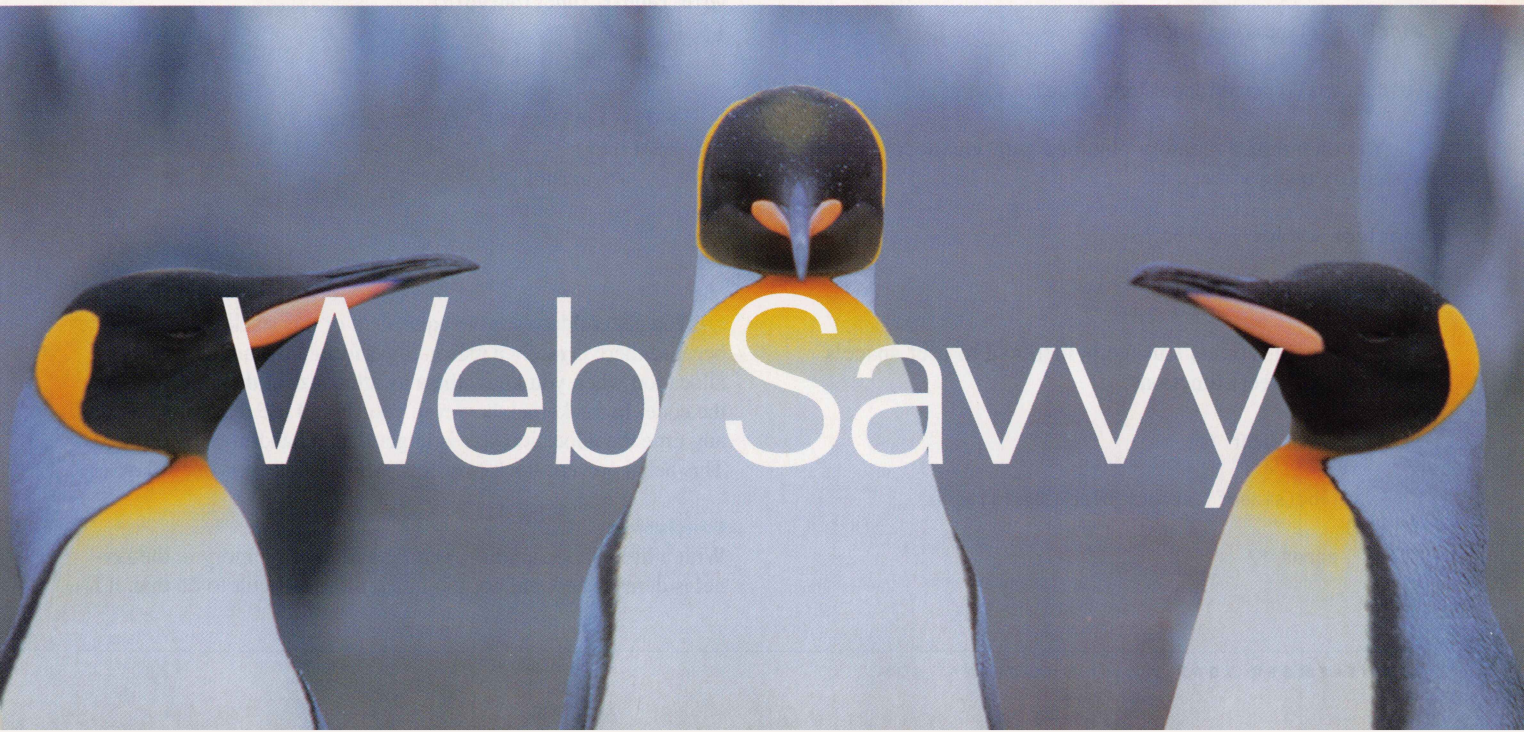
The right management software allows you to seamlessly integrate Linux into your infrastructure.

Linux Management Solutions

No matter what type of Linux implementation you have planned, you'll still need management software to ensure that it's seamlessly integrated across your enterprise. Our management software solutions help maximize the potential of Linux across your entire business by letting you take full advantage of its unique scalability and open-source programming. As a result, you can increase productivity while decreasing total cost of ownership. To find out more about our continuing commitment to Linux innovations, go to ca.com/linux today.

ca Computer Associates®

© 2003 Computer Associates International, Inc. (CA). All rights reserved.



block and return the memory to the kernel:

```
kfree(p);
```

As with free() in user space, calling kfree() on a block of memory that already has been freed or on a pointer that is not an address returned from kmalloc() is a bug, and it can result in memory corruption. Always balance allocations and frees to ensure that kfree() is called exactly once on the correct pointer. Calling kfree() on NULL is checked for explicitly and is safe, although it is not necessarily a sensible idea.

Let's look at the full allocation and freeing cycle:

```
struct sausage *s;

s = kmalloc(sizeof (struct sausage), GFP_KERNEL);
if (!s)
    return -ENOMEM;
/* ... */

kfree(s);
```

Allocating from Virtual Memory

The kmalloc() function returns physically and therefore virtually contiguous memory. This is a contrast to user space's malloc() function, which returns virtually but not necessarily physically contiguous memory. Physically contiguous memory has two primary benefits. First, many hardware devices cannot address virtual memory. Therefore, in order for them to be able to access a block of memory, the block must exist as a physically contiguous chunk of memory. Second, a physically contiguous block of memory can use a single large page mapping. This minimizes the translation lookaside buffer (TLB) overhead of addressing the memory, as only a single TLB entry is required.

Allocating physically contiguous memory has one downside: it is often hard to find physically contiguous blocks of memory, especially for large allocations. Allocating memory that is only virtually contiguous has a much larger chance of success. If you do not need physically contiguous memory, use vmalloc():

```
#include <linux/vmalloc.h>

void * vmalloc(unsigned long size);
```

You then return memory obtained with vmalloc() to the system by using vfree():

```
#include <linux/vmalloc.h>

void vfree(void *addr);
```

Here again, vfree()'s usage is identical to user space's malloc() and free() functions:

```
struct black_bear *p;

p = vmalloc(sizeof (struct black_bear));
if (!p)
    /* error */
/* ... */

vfree(p);
```

In this particular case, vmalloc() might sleep.

Many allocations in the kernel can use vmalloc(), because few allocations need to appear contiguous to hardware devices. If you are allocating memory that only software accesses, such as data associated with a user process, there is no need for the memory to be physically contiguous. Nonetheless, few allocations in the kernel use vmalloc(). Most choose to use kmalloc(), even if it's not needed, partly for historical and partly for performance reasons. Because the TLB overhead for physically contiguous pages is reduced greatly, the performance gains often are well appreciated. Despite this, if you need to allocate tens of megabytes of memory in the kernel, vmalloc() is your best option.

A Small Fixed-Size Stack

Unlike user-space processes, code executing in the kernel has neither a large nor a dynamically growing stack. Instead, each process in the kernel has a small fixed-size stack. The exact size of the stack is architecture-dependent. Most architectures allocate two pages for the stack, so the stack is 8KB on 32-bit machines.

Because of the small stack, allocations that are large, automatic and on-the-stack are discouraged. Indeed, you never should see anything such as this in kernel code:

```
#define BUF_LEN  2048

void rabbit_function(void)
{
    char buf[BUF_LEN];
    /* ...  */
}
```

Instead, the following is preferred:

```
#define BUF_LEN  2048

void rabbit_function(void)
{
    char *buf;

    buf = kmalloc(BUF_LEN, GFP_KERNEL);
    if (!buf)
        /* error! */

/* ... */
}
```

You also seldom see the equivalent of this stack in user space, because there is rarely a reason to perform a dynamic memory allocation when you know the allocation size at the time you write the code. In the kernel, however, you should use dynamic memory any time the allocation size is larger than a handful of bytes or so. This helps prevent stack overflow, which ruins everyone's day.

Conclusion

With a little understanding, getting a hold of memory in the kernel is demystified and not too much more difficult to do than it is

in user space. A few simple rules of thumb can go a long way:

- Decide whether you can sleep (that is, whether the call to kmalloc() can block). If you are in an interrupt handler, in a bottom half, or if you hold a lock, you cannot. If you are in process context and do not hold a lock, you probably can.
- If you can sleep, specify GFP_KERNEL.
- If you cannot sleep, specify GFP_ATOMIC.
- If you need DMA-capable memory (for example, for an ISA or broken PCI device), specify GFP_DMA.
- Always check for and handle a NULL return value from kmalloc().
- Do not leak memory; make sure you call kfree() somewhere.
- Ensure that you do not race and call kfree() multiple times and that you never access a block of memory after you free it.

Robert Love (rml@tech9.net) is a kernel hacker at MontaVista Software and a student at the University of Florida. He is the author of *Linux Kernel Development*. Robert enjoys fine wine and lives in Gainesville, Florida.

RESOURCES

For more information, check out these files in your kernel source tree.

- include/linux/gfp.h: home of the allocation flags.
- include/linux/slab.h: definitions of kmalloc(), et al.
- mm/page_alloc.c: page allocation functions.
- mm/slab.c: implementation of kmalloc(), et al.

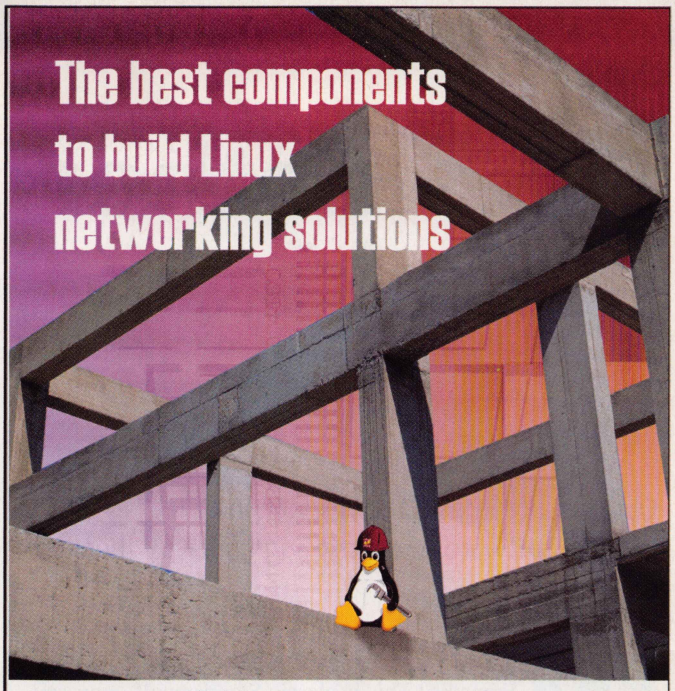
TECH TIP

If you want to run one and only one instance of a console-based program, you can use an alias for the program's name, plus the always-handy GNU screen, which we covered in the January 2003 issue of *Linux Journal*. For example, this alias for mutt detatches the existing screen session named mutt and attaches to it, or if there is no mutt session running, it starts one.

```
alias mutt='screen -d -r mutt || screen -S mutt /usr/bin/mutt'
```

Put it in your .bashrc and try it from two different xterms to see what's going on.

Screen home page: **www.gnu.org/software/screen/screen.html**.



Cyclades communication cards provide all the functionality you need to build Linux®-based network servers, such as routers, console access and remote access servers. Our innovative design and flexibility allow you to build solutions with cost and management advantages without sacrificing performance. By using standard interfaces to communicate with the Operating System, Cyclades makes the integration of your final solution effortless

- Suitable for Remote Access/ Telecommuting, CTI and fax servers
- DSP-based digital V.90 modems

- Suitable for routing, inter-office connectivity and synchronous communications
- One or two WAN interfaces per host card (RS-232/V.35, X.21, T1/E1, G.703)

- Suitable for Console Access, automation and serial applications
- Wide range of models, from 4 up to 64 RS-232 ports per host card

cyclades
Everywhere with Linux

www.cyclades.com/ljb
sales@cyclades.com
1.888.CYCLADES

©2002 Cyclades Corporation. All rights reserved. All other trademarks and product images are property of their respective owners. Product information subject to change without notice.

Put Another Nickel in...

Our chef prepares a tasty selection of featureful jukebox software. *A chacun son goût*—try them all.

BY MARCEL GAGNÉ

What do you mean, you've never heard this song, François? This is classic Steppenwolf, *mon ami*. Bo-rn to be wi-i-ild! What's that? You think I should keep my singing to the shower? I think, François, that you are forgetting who the Chef and owner of this restaurant is. *Non, mon ami*, that was purely informational, and I already have chosen to ignore that subtle dig. Look here, though. Each of these programs allows us to provide customized playlists of our favorite music for our guests.

Speaking of our guests, they are here. Welcome, *mes amis*, to *Chez Marcel*, home of fine Linux fare, exceptional wines and the finest in musical entertainment. Entertainment is, after all, the theme of this issue. Why musical entertainment, you ask? In a much earlier part of my life, I used to make some extra money babysitting some of my parents' friends' children. One of their friends repaired jukeboxes for a living. Consequently, a real honest-to-goodness jukebox was always in the house, full of 45 RPM singles. From time to time, the model and type of machine would change, but with the coin mechanism disabled, I had all the music I wanted available. Now, that's entertainment!

Like a lot of readers out there, I keep a lot of songs on my disk in Ogg Vorbis or MP3 format. This makes sense because although I have a large CD collection, popping CDs into my notebook means carrying the things around, and that's not convenient. So I rip songs from all those CDs I've bought and store them in the tiniest space possible, namely my notebook. The CDs in jewel cases disappear into that virtual space that is my disk, but the system still is far from perfect. Now I have all these songs in a number of large folders with little or no organization. If I want to play something, I have to go searching. Well, no more, *mes amis*! On today's menu, I offer you some great open-source jukebox programs. With any of these on your Linux system, your song collections can become well organized, searchable and fun to use—just like the machines of my youth, only much more compact.

François! What are you still doing here? To the wine cellar, *immédiatement*! I think we need something light today—something fresh and exciting. The 2001 Riesling Auslese Gold Cap Mosel-Saar-Ruwer Wehlener from Germany certainly qualifies as stellar entertainment for that most important of the senses, taste. *Vite*, François!

While François brings back the wine, let's start our tour of jukebox programs available for your Linux system. If you are a KDE desktop user (and even if you are not), you should consider having a look at Scott Wheeler's **JuK** (Figure 1). In terms of pronunciation, think jukebox and you have the idea. This program, being a KDE application, integrates nicely into the KDE desktop, with a tray icon to drop the application out of sight quickly. It includes support for your MP3 and Ogg Vorbis files, collection and playlist management, tag editing and much more.

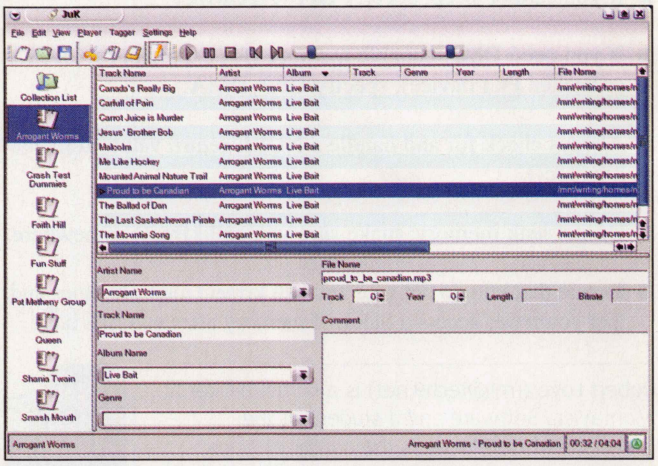


Figure 1. JuK is slated to be part of KDE 3.2.

JuK will be released as part of KDE 3.2 and should be part of any future KDE installation. In the meantime, you can get your copy from the JuK Web site at www.slackorama.net/cgi-bin/content.pl?juk. Building Juk is a simple matter of using the old extract and build five-step:

```
tar -xzvf juk-1.95.tar.gz
cd juk-1.95
./configure
make
su -c "make install"
```

When JuK starts up (with the command `juk`), you find yourself looking at a simple two-pane interface with a sidebar on the left-hand side and a large work/display window on the right, both of which you can size to suit. The expected menu and quick access icons are along the top. The sidebar has a single folder icon labeled Collection List. Adding songs to your JuK box is as simple as clicking the open folder icon, pointing to a folder and selecting the songs you want. Click OK and your songs appear in the default collection list. Because organization is key, you can create additional playlists, then drag and drop titles into your folder of choice. The titles appear in your new playlist but remain in the master collection as well.

From the main track list window, you can click each of the columns to sort by album title, artist, date or other catalog identifier. As anyone with a collection of songs on their computer knows, the information contained in the information tags isn't always perfect. To deal with this problem, we have JuK's power tag editor. What I really like about this feature is you can edit information in-line by right-clicking on a song title, as well as by bringing up the more comprehensive tag editor.

sgi
The universe will have fewer secrets. sgi.com/altix
The SGI® Altix™ 3000 family of servers and superclusters, the Intel® Itanium® 2 processor, and the Linux® operating system have teamed to bring you high-performance, reliable, and cost-effective servers. SGI has been the power behind some of the most dramatic breakthroughs in science, manufacturing, media, energy exploration, and government. Let SGI power your complex and data-intensive deployments and help you discover exactly what makes the universe tick.
SGI. THE SOURCE OF INNOVATION AND DISCOVERY™
intel inside
ITANIUM 2
©2003 Silicon Graphics, SGI, and the SGI logo are registered trademarks and Altix and The Source of Innovation and Discovery are trademarks of Silicon Graphics, Inc. in the United States and/or other countries worldwide. Intel, Itanium, Intel Inside, and the Intel Inside logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. Linux is a registered trademark of Linus Torvalds, used with permission by Silicon Graphics, Inc.

Let's say you have 12 songs from a single album and you'd like to enter the date for that album, JuK's editor lets you select multiple titles and do mass changes to selected fields. This is extremely handy and a great time-saver.

On the GNOME side of things, allow me to introduce you to Colin Walters' **Rhythmbox** (Figure 2). Rhythmbox is another slick-looking jukebox program. You can search for songs, sort by a variety of fields, create playlists and so on. The music library browser views can be configured for two or three columns listing artist, genre or album title. Click on any of the choices, and Rhythmbox narrows down your choices. There's a tag editor for your collection but no in-line edit feature (at least, not at the time of this writing). It even has built-in support for Internet Radio; with a tuner card, you can set your favorite stations and listen from the same interface. You can pick up your copy of Rhythmbox at www.rhythmbox.org.

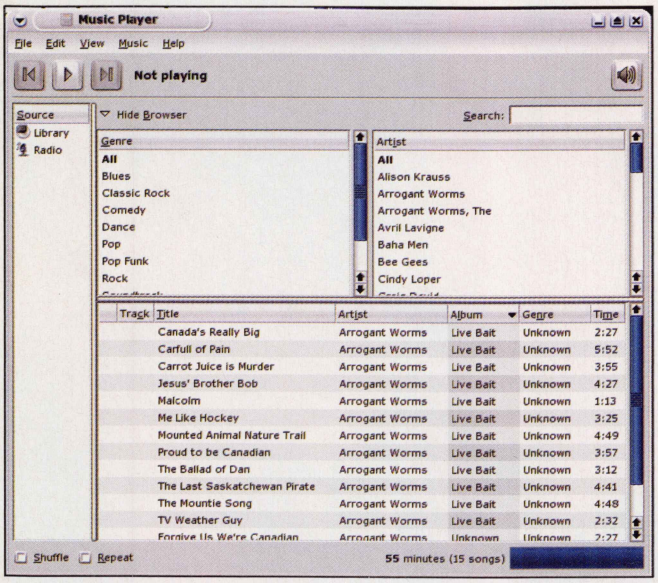


Figure 2. Your Very Own Rhythmbox

Rhythmbox plays its songs using the gstreamer package, so that's an important prerequisite. Building Rhythmbox from source isn't difficult, but I did wind up spending some time looking for all the various required GNOME development packages. These included the development packages for bonobo, GTK+, pango, gnomeui, atk, gconf and others. I mention this because I found myself going back quite a few times for packages I was missing when building this one. Binaries are available for those who would like a shortcut. For a great RPM repository, try rpm.pbone.net. But I digress; Rhythmbox is another extract and build five-step project, so no big surprises there:

```
tar -xzvf rhythmbox-0.5.3.tar.gz
cd rhythmbox-0.5.3
./configure
make
su -c "make install"
```

When you start Rhythmbox for the first time, you're met with a small question-and-answer dialog. As part of that setup, you are asked where on the system you store your music. You can either provide a pathname here or skip the step entirely. If you skip it, you can add your songs manually. Rhythmbox reads the song titles, artist name, album title and so on from each cut's tag and sorts the whole thing automatically. Using the tag editor, you also can rate the songs from zero to five stars and then use that rating to look for your favorite tunes.

The final item on tonight's menu bills itself as "the one music manager to rule them all", a boast that certainly makes you take notice. On that note, I think we all should take a minute to sip our wines. The program to rule them all is Andreas Klöckner's **Madman** (Figure 3), and it is housed at madman.sourceforge.net/index.php.

Should you decide to build Madman from source, you need to do a reprise of your old friend, the extract and build five-step. At the time of this writing, Madman's source release package was madman-0.91.1.tar.gz.

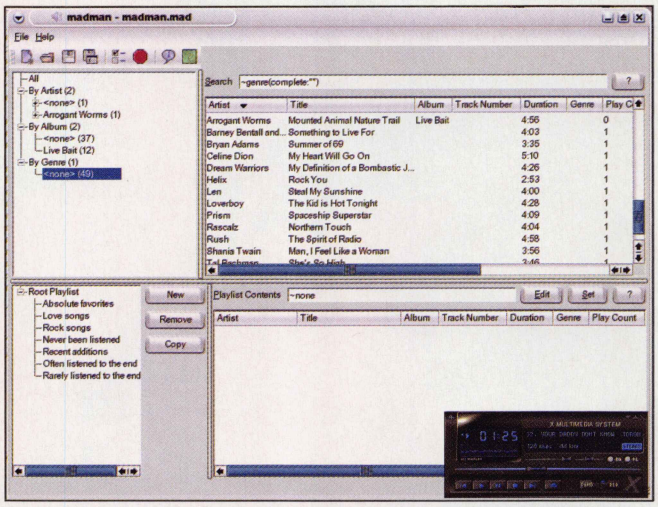


Figure 3. You don't have to be mad to use Madman.

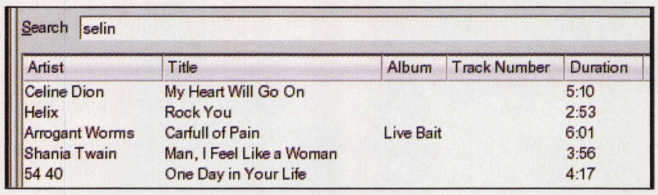


Figure 4. Madman lets you search for a song even if you can't spell it.

Obviously, the basic features of Madman are much the same as JuK's and Rhythmbox's in that it provides an interface that allows you to organize your music collection and play selected songs. That said, Madman really is different—it even feels different. For instance, the search function is ridiculously flexible. Let's say you wanted to play a song by Celine Dion, but you didn't know how to spell Celine. Simply enter `selin` or something that looks like it might be right, and Madman locates the song for you (Figure 4). *Mais oui*, of course I know how to spell it.

But wait, there's more. In the bottom part of Madman's main screen sits another one of Madman's cool features. The program automatically tracks which songs you've listened to completely and how often. So, you create an automatic list of

favorites, simply by listening. You also can right-click on a song and assign it a rating using a five-star system.

Madman starts by asking you to enter directories in which you store your songs. This is a simple add dialog where you can add as many as you like. All the music in your collection can be accessed by artist, genre or album title. You also can search for a song using Madman's aforementioned super-flexible search engine. When you find a song you like, highlight it and press F9; press F10 to play it next. Alternatively, you can choose totally random select.

While the other two programs I looked at tonight play directly from the interface, with the machinery hidden in the background, Madman fires up XMMS to do the actual playing. An option can be found in the preferences dialog to select a different player, but at this time, XMMS seems to be the only option available.

There's a lot to like about Madman. When I rip and encode CDs, I sometimes find the information isn't properly carried through into the tag. As a result, I'll get a blank listing. No problem; press F9 to play it, check the title in XMMS's display and edit the tag in-line on Madman's display with a slow double-click.

What did you say, François? Ah, thank you. I'm sorry, *mes amis*, but it appears as though we have sailed right through another evening. Where does the time go, eh? Still, at *Chez Marcel*, we never want you to leave without one more leisurely glass of wine, so I will have François pour you a final glass. Since you all have Linux systems at your tables, why not bring up the jukebox program of your choice, put on the headphones and listen to one last song while you finish your wine. Until next time, *mes amis*, let us drink to one another's health. *A vôtre santé! Bon appétit!*

Marcel Gagné (mggagne@salmar.com) lives in Mississauga, Ontario. He is the author of the newly published *Moving to Linux: Kiss the Blue Screen of Death Goodbye!* (ISBN 0-321-15998-5) from Addison Wesley. His first book is the highly acclaimed *Linux System Administration: A User's Guide* (ISBN 0-201-71934-7). In real life, he is president of Salmar Consulting, Inc., a systems integration and network consulting firm.

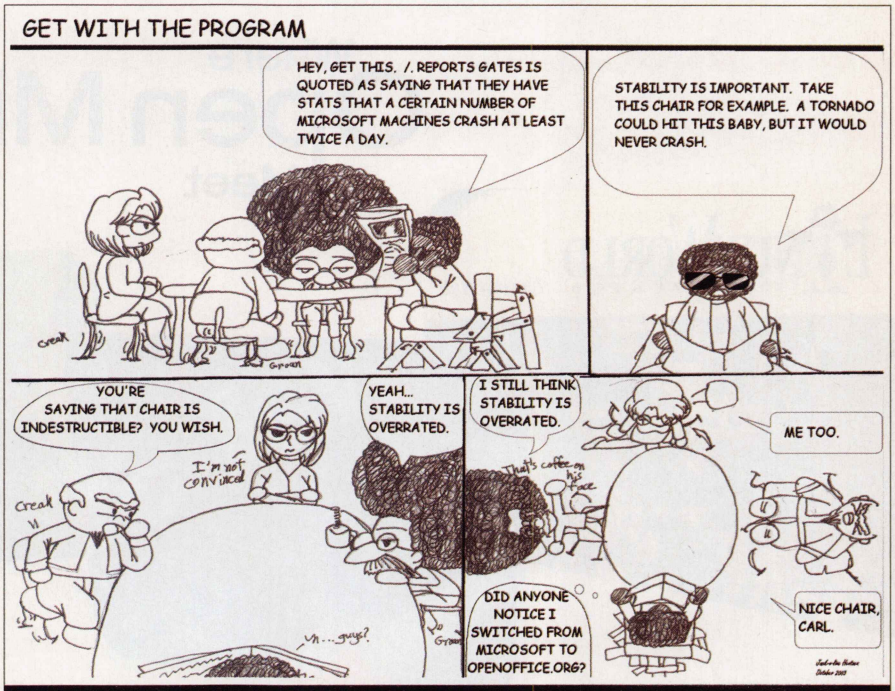
RESOURCES

JuK: www.slackorama.net/cgi-bin/content.pl?juk

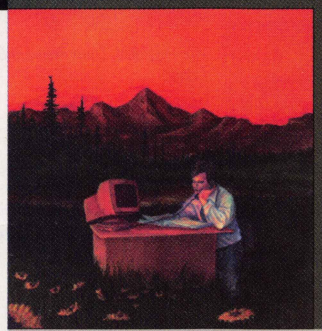
Madman: madman.sourceforge.net

Rhythmbox: www.rhythmbox.org

Marcel's Wine Page: www.marcelgagne.com/wine.html



OOoSwitch:
501 Things You Wanted To Know About Switching to OpenOffice.org from Microsoft® Office



Hentzenwerke Publishing

ISBN 1-930919-36-0 312 pages $49.95

Moving to OpenOffice.org?

OOoSwitch will help you become productive with OOo 1.1 quickly.

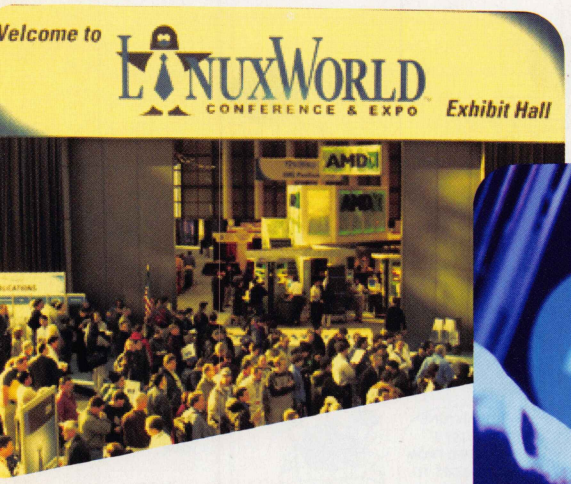
Download free sample chapters and order online at www.hentzenwerke.com

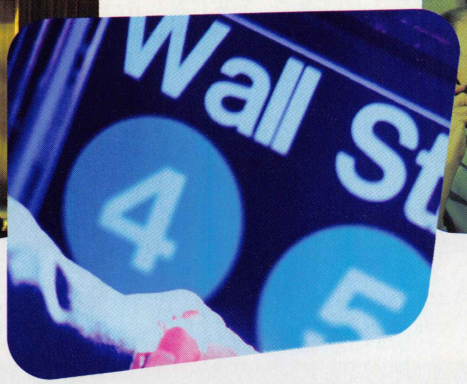
Also available at Microcenter and Barnes & Noble stores nationwide

Hentzenwerke

LINUXWORLD® CONFERENCE & EXPO

Where Open Minds Meet





Keynote Speakers

Chris Stone
Vice Chairman, Office of the CEO, Novell, Inc.
Wednesday, January 21
9:15 am – 10:00 am

Dave Dargo
Vice President, Linux Program Office, Oracle
Wednesday, January 21
11:45 am – 12:30 pm

Tom Killalea
Vice President of Infrastructure, Amazon.com
Wednesday, January 21
2:45 pm – 3:30 pm

Sam Greenblatt
Senior Vice President and Chief Architect, Computer Associates Intl.
Thursday, January 22
9:15 am – 10:00 am

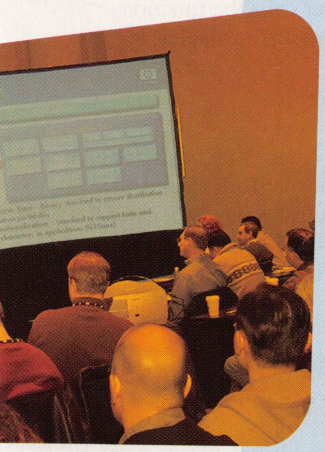
Ross A. Mauri
General Manager, e-Business on demand, IBM Systems Group
Thursday, January 22
1:45 pm – 2:30 pm

CONFERENCE: January 20 – 23, 2004

EXPO: January 21 – 23, 2004

THE JAVITS CENTER: New York, NY

LinuxWorld Conference & Expo is the world's leading and most comprehensive event focusing on Linux and open source solutions. This January, discover how companies across the globe have achieved higher profits and increased their productivity by utilizing Linux, the fastest growing operating system in the world.



- Get in-depth coverage on the latest Linux and open source developments and learn about innovative product solutions from the industry's top companies.
- Network with the leaders in the open source movement and discuss with your peers how to best leverage the technology for your organization.
- Participate in LinuxWorld's world-class education program and benefit from interactive training in the all-new Hands-on Labs!
- Attend the 2nd annual Linux Financial Summit sponsored by BEA, and hear how Wall Street firms are leveraging Linux to achieve a lower total cost of ownership.
- Discover real world practical solutions and find answers to today's most critical IT issues.

Cornerstone Sponsor

ORACLE

Platinum Sponsors

ca Computer Associates · IBM · intel · hp invent · Novell · Sun microsystems · SUSE

REGISTER ONLINE WITH REFERENCE CODE: A-LJD

www.linuxworldexpo.com

IDG WORLD EXPO

I2C Drivers, Part I

The I2C bus helps you monitor the health of your system. Here's how to develop a driver that will get you all the hardware info you need to know.

BY GREG KROAH-HARTMAN

In the June and August 2003 issues of *Linux Journal*, my column covered the Linux kernel driver model, and the I2C subsystem was used as an example. This month, we discuss what the I2C subsystem does and how to write a driver for it.

I2C is the name for a two-wire serial bus protocol originally developed by Phillips. It commonly is used in embedded systems so different components can communicate; PC motherboards use I2C to talk to different sensor chips. Those sensors typically report back fan speeds, processor temperatures and a whole raft of system hardware information. The protocol also is used in some RAM chips to report information about the DIMM itself back to the operating system.

The I2C kernel code has lived outside of the main kernel tree for much of its development life—it originally was written back in the 2.0 days. The 2.4 kernel contains a bit of I2C support, mainly for some video drivers. With the 2.6 kernel, a large portion of the I2C code has made it into the main kernel tree, thanks to the effort of a number of kernel developers who changed the interfaces to be more acceptable to the kernel community. A few drivers still live only in the external CVS tree and have not been moved into the main kernel.org tree, but it is only a matter of time before they, too, are ported.

The I2C kernel code is broken up into a number of logical pieces: the I2C core, I2C bus drivers, I2C algorithm drivers and I2C chip drivers. We ignore how the I2C core operates in this article and focus instead on how to write a bus and algorithm driver. In Part II, we will cover how to write an I2C chip driver.

I2C Bus Drivers

An I2C bus driver is described by a struct named i2c_adapter, which is defined in the include/linux/i2c.h file. Only the following fields need to be set up by the bus driver:

- struct module *owner; —set to the value (THIS_MODULE) that allows the proper module reference counting.
- unsigned int class; —the type of I2C class devices that this driver supports. Usually this is set to the value I2C_ADAP_CLASS_SMBUS.
- struct i2c_algorithm *algo; —a pointer to the struct i2c_algorithm structure that describes the way data is transferred through this I2C bus controller. More information on this structure is provided below.
- char name[I2C_NAME_SIZE]; —set to a descriptive name of the I2C bus driver. This value shows up in the sysfs filename associated with this I2C adapter.

The code below comes from an example I2C adapter driver called tiny_i2c_adap.c, available from the *Linux Journal* FTP site [ftp.ssc.com/pub/lj/listings/issue116/7136.tgz] and shows how the struct i2c_adapter is set up:

```
static struct i2c_adapter tiny_adapter = {
    .owner   = THIS_MODULE,
    .class   = I2C_ADAP_CLASS_SMBUS,
    .algo    = &tiny_algorithm,
    .name    = "tiny adapter",
};
```

To register this I2C adapter, the driver calls the function i2c_add_adapter with a pointer to the struct i2c_adapter:

```
retval = i2c_add_adapter(&tiny_adapter);
```

If the I2C adapter lives on a type of device that has a struct device associated with it, such as a PCI or USB device, then before the call to i2c_add_adapter, the adapter device's parent pointer should be set to that device. This pointer configuration can be seen in the following line from the drivers/i2c/busses/i2c-piix4.c driver:

```
/* set up sysfs linkage to our parent device */
piix4_adapter.dev.parent = &dev->dev;
```

If this parent pointer is not set up, the I2C adapter is positioned on the legacy bus and shows up in the sysfs tree at /sys/devices/legacy. Here is what happens to our example driver when it is registered:

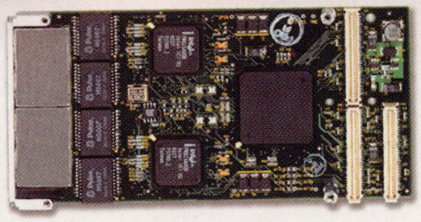
```
$ tree /sys/devices/legacy/
/sys/devices/legacy/
|-- detach_state
|-- floppy0
|   |-- detach_state
|   `-- power
|       `-- state
|-- i2c-0
|   |-- detach_state
|   |-- name
|   `-- power
|       `-- state
`-- power
    `-- state
```

As discussed in the previous kernel driver model columns, the I2C adapter also shows up in the /sys/class/i2c-adapter directory:

Linux On Demand

You Decide...Linux ready on every board

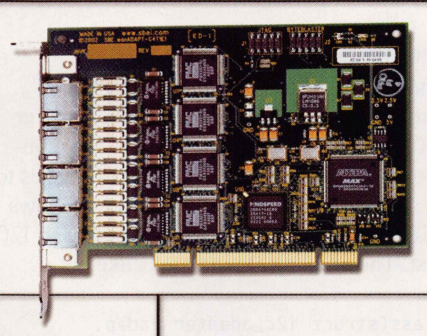
lanPMC-4Gx



Quad port Gigabit Ethernet copper or fiber PMC adapter

- 2 or 4 auto-negotiating 10/100/1000 Ethernet ports
- Fiber version supports long and short reach
- Auto sensing speed, simplex/duplex, flow control, and detection of polarity and cable lengths on copper version

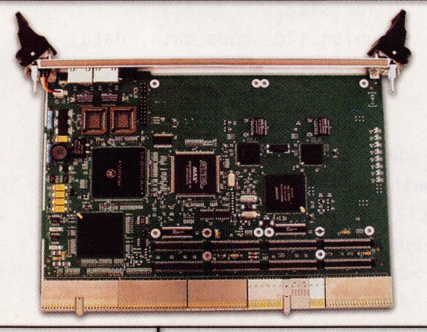
wanADAPT-C4T1E1



Channelized quad port T1/E1/J1 PCI WAN adapter

- 128 channels (DS0)
- Line buildouts from DSX to DS1
- Supports D4, SF, ESF, SLC-96, JT-6704 framing
- Selectable clock options

HW400c/M DKL



6U CPCI processing blade with dual PTMC sites

- 333 MHz PowerPC CPU
- PICMG 2.16 Packet Switching Backplane enabled
- Dual 10/100 Ethernet ports
- Includes TimeSys Linux GPL Software Developer's Kit
- Drivers for T1/E1, T3/E3, HSSI, V.35, X.21 & other SBE PMCs

SBE®

WAN • LAN • Storage • Carriers • Custom

flexibility on demand | 925-355-2000 | info@sbei.com | www.sbei.com

```
$ tree /sys/class/i2c-adapter/
/sys/class/i2c-adapter/
`-- i2c-0
    |-- device -> ../../../devices/legacy/i2c-0
    `-- driver -> ../../../bus/i2c/drivers/i2c_adapter
```

To unregister an I2C adapter, the driver should call the function i2c_del_adapter with a pointer to the struct i2c_adapter, like this:

```
i2c_del_adapter(&tiny_adapter);
```

I2C Algorithm Drivers

An I2C algorithm is used by the I2C bus driver to talk to the I2C bus. Most I2C bus drivers define their own I2C algorithms and use them, as they are tied closely to how the bus driver talks to that specific type of hardware. For some classes of I2C bus drivers, a number of I2C algorithm drivers already have been written. Examples of these are ITE adapters found in drivers/i2c/i2c-algo-ite.c, IBM PPC 405 adapters found in drivers/i2c/i2c-algo-ibm_ocp.c and a generic I2C bit shift algorithm found in drivers/i2c/i2c-algo-bit.c. All of these already written algorithms have their own functions with which an I2C bus driver needs to register to use. For more information on these, please see all of the drivers/i2c/i2c-algo-*.c files in the kernel tree.

For our example driver, we are going to create our own I2C algorithm driver. An algorithm driver is defined by a struct i2c_algorithm structure and is defined in the include/linux/i2c.h file. Here is a description of some of the commonly used fields:

- char name[32];: the name of the algorithm.
- unsigned int id;: description of the type of algorithm this structure defines. These different types are defined in the include/linux/i2c-id.h file and start with the characters I2C_ALGO_.
- int (*master_xfer)(struct i2c_adapter *adap,struct i2c_msg msgs[], int num);: a function pointer to be set if this algorithm driver can do I2C direct-level accesses. If it is set, this function is called whenever an I2C chip driver wants to communicate with the chip device. If it is set to NULL, the smbus_xfer function is used instead.
- int (*smbus_xfer) (struct i2c_adapter *adap, u16 addr, unsigned short flags, char read_write, u8 command, int size, union i2c_smbus_data *data);: a function pointer to be set if this algorithm driver can do SMB bus accesses. Most PCI-based I2C bus drivers are able to do this, and they should set this function pointer. If it is set, this function is called whenever an I2C chip driver wants to communicate with the chip device. If it is set to NULL, the master_xfer function is used instead.
- u32 (*functionality) (struct i2c_adapter *);: a function pointer called by the I2C core to determine what kind of reads and writes the I2C adapter driver can do.

In our example I2C adapter driver, the i2c_adapter structure referenced the tiny_algorithm variable. That structure is defined as the following:

```
static struct i2c_algorithm tiny_algorithm = {
    .name           = "tiny algorithm",
    .id             = I2C_ALGO_SMBUS,
    .smbus_xfer     = tiny_access,
    .functionality  = tiny_func,
};
```

The tiny_func function is small and tells the I2C core what types of I2C messages this algorithm can support. For this driver, we want to be able to support a few different I2C message types:

```
static u32 tiny_func(struct i2c_adapter *adapter)
{
    return I2C_FUNC_SMBUS_QUICK |
           I2C_FUNC_SMBUS_BYTE |
           I2C_FUNC_SMBUS_BYTE_DATA |
           I2C_FUNC_SMBUS_WORD_DATA |
           I2C_FUNC_SMBUS_BLOCK_DATA;
}
```

All of the different I2C message types are defined in include/linux/i2c.h and start with the characters I2C_FUNC_.

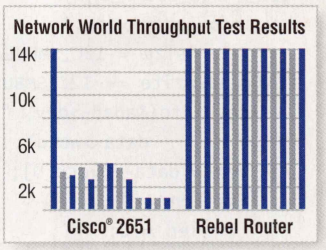
The tiny_access function is called when an I2C client driver wants to talk to the I2C bus. Our example function is quite simple; it merely logs all of the requests the I2C chip driver makes to the syslog and reports success back to the caller. This log allows you to see all of the different addresses and data types that an I2C chip driver may request. The implementation looks like:

```
static s32 tiny_access(struct i2c_adapter *adap,
                       u16 addr,
                       unsigned short flags,
                       char read_write,
                       u8 command,
                       int size,
                       union i2c_smbus_data *data)
{
    int i, len;

    dev_info(&adap->dev, "%s was called with the "
             "following parameters:\n",
             __FUNCTION__);
    dev_info(&adap->dev, "addr = %.4x\n", addr);
    dev_info(&adap->dev, "flags = %.4x\n", flags);
    dev_info(&adap->dev, "read_write = %s\n",
             read_write == I2C_SMBUS_WRITE ?
             "write" : "read");
    dev_info(&adap->dev, "command = %d\n",
             command);

    switch (size) {
    case I2C_SMBUS_PROC_CALL:
        dev_info(&adap->dev,
                 "size = I2C_SMBUS_PROC_CALL\n");
        break;
```

THE

REVOLUTION RELOADED

What is router performance? Is router performance defined by the biggest brand, or can you experience it in the real world? It's time to get some answers when you hear a little voice inside asking, **"What did I get for the money?"** Many have paid a high price to own the biggest brand, but can they accept the truth? In a recent head-to-head lab test conducted by Network World Magazine (Vol 20, Issue 28), the industry's largest router manufacturer declined to participate. When asked why, they said, **"... users are interested in things other than performance."** Can you imagine them saying this to their customers? In the past five years, ImageStream has introduced the world's lowest cost routers for OC12/STM4, OC3/STM1, DS3/E3 and high-density T1/E1 applications. Now, independent tests prove again that **ImageStream routers blow away the monster brand** in terms of throughput, latency, and jitter. And if the best price and performance weren't enough, every ImageStream router includes a 31-day performance guarantee, 24/7 technical support, and free software upgrades for life. Linux® has always made the impossible possible. **The choice is yours, the red LED or the blue?**



Independent tests confirm ImageStream's superior performance in terms of throughput, latency and jitter.

ImageStream™
Routers for the Real World

800.813.5123
www.imagestream.com
1.574.935.8484

© Copyright 2003 ImageStream Internet Solutions, Inc. All rights reserved worldwide. ImageStream and "Routers for the Real World" are trademarks of ImageStream Internet Solutions, Inc. Linux is a registered trademark of Linus Torvalds. All other trademarks are property of their respective owners.

```
    case I2C_SMBUS_QUICK:
        dev_info(&adap->dev,
                 "size = I2C_SMBUS_QUICK\n");
        break;
    case I2C_SMBUS_BYTE:
        dev_info(&adap->dev,
                 "size = I2C_SMBUS_BYTE\n");
        break;
    case I2C_SMBUS_BYTE_DATA:
        dev_info(&adap->dev,
                 "size = I2C_SMBUS_BYTE_DATA\n");
        if (read_write == I2C_SMBUS_WRITE)
            dev_info(&adap->dev,
                     "data = %.2x\n", data->byte);
        break;
    case I2C_SMBUS_WORD_DATA:
        dev_info(&adap->dev,
                 "size = I2C_SMBUS_WORD_DATA\n");
        if (read_write == I2C_SMBUS_WRITE)
            dev_info(&adap->dev,
                     "data = %.4x\n", data->word);
        break;
    case I2C_SMBUS_BLOCK_DATA:
        dev_info(&adap->dev,
                 "size = I2C_SMBUS_BLOCK_DATA\n");
        if (read_write == I2C_SMBUS_WRITE) {
            dev_info(&adap->dev, "data = %.4x\n",
                     data->word);
            len = data->block[0];
            if (len < 0)
                len = 0;
            if (len > 32)
                len = 32;
            for (i = 1; i <= len; i++)
                dev_info(&adap->dev,
                         "data->block[%d] = %x\n",
                         i, data->block[i]);
        }
        break;
    }
    return 0;
}
```

Now that the tiny_i2c_adap driver is built and loaded, what can it do? On its own, it cannot do anything. An I2C bus driver needs an I2C client driver in order to do anything besides sit in the sysfs tree. So, if the lm75 I2C client driver is loaded, it tries to use the tiny_i2c_adap driver to find the chip for which it was written:

```
$ modprobe lm75
$ tree /sys/bus/i2c/
/sys/bus/i2c/
|-- devices
|   |-- 0-0048 -> ../../../devices/legacy/i2c-0/0-0048
|   |-- 0-0049 -> ../../../devices/legacy/i2c-0/0-0049
|   |-- 0-004a -> ../../../devices/legacy/i2c-0/0-004a
|   |-- 0-004b -> ../../../devices/legacy/i2c-0/0-004b
|   |-- 0-004c -> ../../../devices/legacy/i2c-0/0-004c
|   |-- 0-004d -> ../../../devices/legacy/i2c-0/0-004d
|   |-- 0-004e -> ../../../devices/legacy/i2c-0/0-004e
|   `-- 0-004f -> ../../../devices/legacy/i2c-0/0-004f
`-- drivers
    |-- i2c_adapter
    `-- lm75
        |-- 0-0048 -> ../../../../devices/legacy/i2c-0/0-0048
        |-- 0-0049 -> ../../../../devices/legacy/i2c-0/0-0049
        |-- 0-004a -> ../../../../devices/legacy/i2c-0/0-004a
        |-- 0-004b -> ../../../../devices/legacy/i2c-0/0-004b
        |-- 0-004c -> ../../../../devices/legacy/i2c-0/0-004c
        |-- 0-004d -> ../../../../devices/legacy/i2c-0/0-004d
        |-- 0-004e -> ../../../../devices/legacy/i2c-0/0-004e
        `-- 0-004f -> ../../../../devices/legacy/i2c-0/0-004f
```

Because the tiny_i2c_adap driver responds with a success to every read and write request it is asked to accomplish, the lm75 I2C chip driver thinks it has found an lm75 chip at every known possible I2C address for this chip. This abundance of addresses is why I2C devices 0-0048 through 0-004f have been created. If we look at the directory for one of these devices, the sensor files for this chip driver are shown:

```
$ tree /sys/devices/legacy/i2c-0/0-0048/
/sys/devices/legacy/i2c-0/0-0048/
|-- detach_state
|-- name
|-- power
|   `-- state
|-- temp_input
|-- temp_max
`-- temp_min
```

The detach_state file and power directory is created by the kernel driver core and is used for power management. It is not created by the lm75 driver. The functions of the other files in this directory are described below.

If we ask the lm75 driver for the current value of temp_max, we receive the following:

```
$ cat /sys/devices/legacy/i2c-0/0-0048/temp_max
1000
```

To get that value, the lm75 driver asked the tiny_i2c_adap driver to read some addresses on the I2C bus. This request is shown in the syslog:

```
$ dmesg
i2c_adapter i2c-0: tiny_access was called with the following parameters:
i2c_adapter i2c-0: addr = 0048
i2c_adapter i2c-0: flags = 0000
i2c_adapter i2c-0: read_write = read
i2c_adapter i2c-0: command = 0
i2c_adapter i2c-0: size = I2C_SMBUS_WORD_DATA
i2c_adapter i2c-0: tiny_access was called with the following parameters:
i2c_adapter i2c-0: addr = 0048
i2c_adapter i2c-0: flags = 0000
i2c_adapter i2c-0: read_write = read
i2c_adapter i2c-0: command = 3
i2c_adapter i2c-0: size = I2C_SMBUS_WORD_DATA
```

The detach_state file and power directory is created by the kernel driver core and is used for power management.

```
i2c_adapter i2c-0: tiny_access was called with
➥the following parameters:
i2c_adapter i2c-0: addr = 0048
i2c_adapter i2c-0: flags = 0000
i2c_adapter i2c-0: read_write = read
i2c_adapter i2c-0: command = 2
i2c_adapter i2c-0: size = I2C_SMBUS_WORD_DATA
```

The log file shows that the tiny_access function was called three times. The first command wanted to read a word of data from register 0 out of the device with the address 0048. The second and third reads asked for register 3 and register 2 from the same device. The commands match up with the following code from the drivers/i2c/chips/lm75.c file in the lm75_update_client function:

```
data->temp_input = lm75_read_value(client,
                                   LM75_REG_TEMP);
data->temp_max = lm75_read_value(client,
                                 LM75_REG_TEMP_OS);
data->temp_hyst = lm75_read_value(client,
                                  LM75_REG_TEMP_HYST);
```

The lm75_read_value function in that same file contains the following code:

```
/* All registers are word-sized, except
for the configuration register. LM75 uses
a high-byte first convention, which is
exactly opposite to the usual practice. */
static int lm75_read_value(struct i2c_client
                           *client, u8 reg)
{
  if (reg == LM75_REG_CONF)
      return i2c_smbus_read_byte_data(client,
                                      reg);
  else
      return swap_bytes(
             i2c_smbus_read_word_data(client,
                                      reg));
}
```

Therefore, when the lm75 driver wants to read the value of the max temperature, it calls the lm75_read_value function with the register number, which then calls the I2C core function i2c_smbus_read_word_data. That I2C core function looks up on which I2C bus the client device is, and then it calls the I2C algorithm associated with that specific I2C bus to do the data transfer. This is the method, then, by which our tiny_i2c_adap driver is asked to complete the transfer.

If this same sysfs file is written to, the lm75 driver asks the tiny_i2c_adap driver to write some data to a specific address on the I2C bus in the same way the read was requested. This request also is shown in the syslog:

```
$ echo 300 > /sys/devices/legacy/i2c-0/
➥0-0048/temp_max
$ dmesg
i2c_adapter i2c-0: tiny_access was called
➥with the following parameters:
i2c_adapter i2c-0: addr = 0048
i2c_adapter i2c-0: flags = 0000
i2c_adapter i2c-0: read_write = write
i2c_adapter i2c-0: command = 3
i2c_adapter i2c-0: size = I2C_SMBUS_WORD_DATA
i2c_adapter i2c-0: data = 8000
```

Conclusion

This month we covered the basics of the I2C driver subsystem and explained how to write a simple I2C bus and I2C algorithm driver that work with any existing I2C client driver. The complete driver, dmn-09-i2c-adap.c, is available from the *Linux Journal* FTP site at ftp.ssc.com/pub/lj/listings/issue116/7136.tgz. In Part II, we will cover how to write an I2C chip driver.

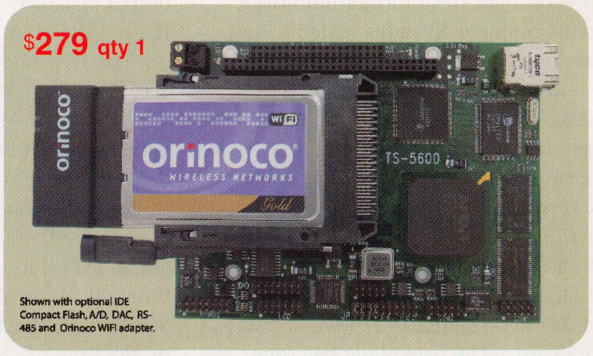
Greg Kroah-Hartman currently is the Linux kernel maintainer for a variety of different driver subsystems. He works for IBM, doing Linux kernel-related things and can be reached at greg@kroah.com.

Wireless Connectivity
- Cellular Modem
- 802.11b WiFi

TS-5600 586 Single Board Computer

Linux 2.4

$279 qty 1



Shown with optional IDE Compact Flash, A/D, DAC, RS-485 and Orinoco WiFi adapter.

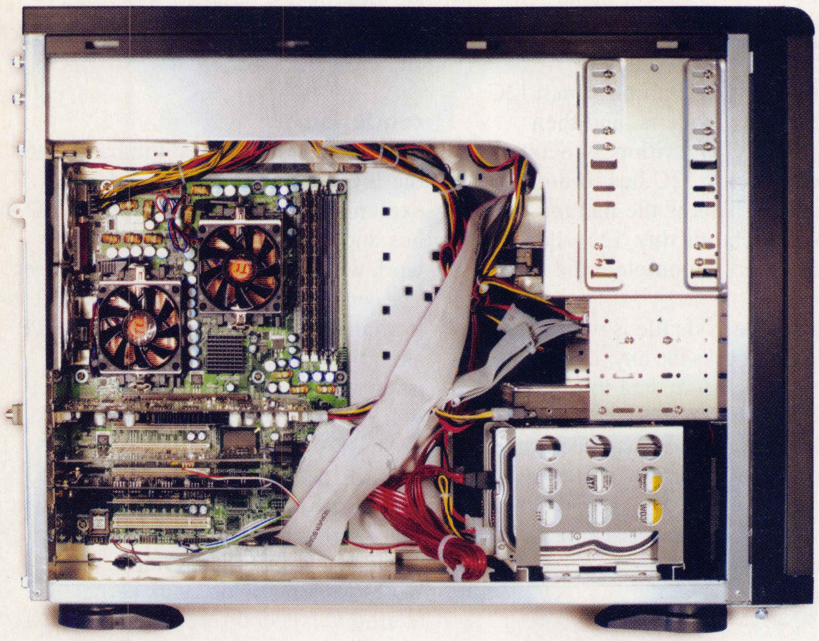
- PC compatible: BIOS, DOS, RTC, I/O & memory map
- User programmable PLD for digital I/O
- Low power +5v @ 900mA
- 4.3" x 5.6"
- TS-Linux 3.0 is open source GNU/Linux - Glibc 2.2.5
- Apache, SSH, Perl, bash, FTP, kernel drivers

- 133 MHz x86
- PCMCIA slot
- 10/100 Ethernet
- PC/104 Bus
- Compact Flash
- 32-64 MB SDRAM
- 6 ch 12-bit A/D
- 2 ch 12-bit DAC
- 23 Digital I/O
- 2 Serial Ports

Technologic SYSTEMS

(480)-837-5200
www.embedded586.com

Ultimate Linux Box



AMD64 looks great on paper, but can you build a high-end system with 64-bit processors and have all your device drivers work? We mix the hottest new processor architecture with the latest video, audio and storage devices and watch for smoke. **BY GLENN STONE**

For the Ultimate Linux Box this year, Editor in Chief Don Marti and I decided to build a multimedia workstation. I was thinking of doing a dual-Xeon unit, having helped build a few of them for a former employer, when Trey Harris of Monarch Computer Systems called to say, "The dual Opteron boards are out." My mouth watered at the thought. The *only* formally released OS that runs on an Opteron (at press time) is Linux. We traded e-mails and machines, and ultimately we came up with a design that, although it has a rough edge or two as I write this, should be ready to go by the time this is printed.

The motherboard is an Arima HDAMB workstation board in an ATX form factor, with the AMD 8111/8151 Rhapsody chipset. It has slots for four DDR333 Registered ECC DIMMs (we used two 1GB Corsair CM74SD1024RLP-2700s), an 8X AGP card and five 32-bit PCI cards. The backplane has four USB jacks, a Broadcom gigabit Ethernet interface and a Realtek ALC650 sound system, as well as the usual parallel, serial and PS/2 keyboard and mouse ports (one of each). The test machine did not come with a FireWire interface or serial ATA ports, although the board does offer options for both. The twin Socket 940s were loaded with AMD Opteron 246s running at 2GHz and topped with Thermaltake AI724 coolers.

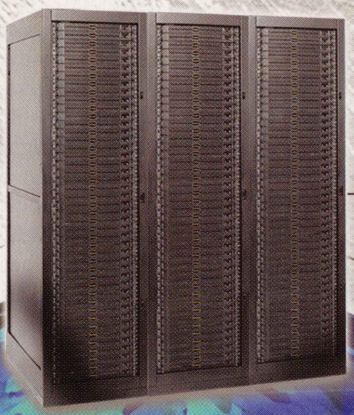
Video is provided by an NVIDIA Quadro FX 1000. This is a workstation-class, twin-head-capable card that I selected because NVIDIA released support for it on the AMD64 back in December 2002. (ATI has not released 64-bit support for its competing product, the Fire GL X1, as of press time.) Sound comes from the Creative Labs Sound Blaster Audigy 2, a 24-bit/192kHz, THX-certified digital card capable of 6.1 analog output or Dolby Digital EX on the digital side. It also has an IEEE 1394 (FireWire) interface.

The IDE interfaces are populated by a JLMS XJ-HD166S 16x DVD/ROM drive and a LITE-ON LTR-52327S 52/32/52x CD-RW drive. We chose to forego a DVD writer, because the whole issue of DVD+RW/DVD-RW/DVD-RAM—and Linux drivers for each—is a rather sticky issue. Part of the joy of Linux and of open standards such as IDE, however, is that experienced audiophiles are free to choose their own favorite and expect it to work when they plug it in. Indeed, at press time, all of the above combination drives are on the market, and both drives and drivers probably will make quantum leaps by the time this issue hits the newsstand.

If we're using the IDE interfaces for optical drives, what about hard disks? Indeed, aside from the Opterons themselves, hard disks are the most formidable part of this system. I was eager to try out 3ware's new Serial ATA (SATA) RAID card, the Escalade 8500-4. It is a four-channel, half-size 64-bit PCI card that also works well in 32-bit slots. At press time, the Escalade 8506s were becoming available, replacing the 8500s. According to 3ware's Product Transition Matrix, the second-generation controller is up to 25% faster than the 8500s. The 8500 is what we had to test with, though, and the statistics reflect that. The good news

HIGH PERFORMANCE COMPUTING

SOLUTIONS FOR YOUR WORLD

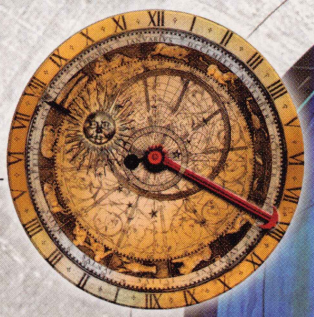


Cluster-tailored Solutions

- Life Sciences (BioBrew)
- Digital Content Creation (RenderBrew)
- Petrochemical (PetroBrew)

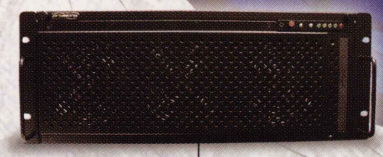
Cluster Management Software

- Callident Rx
- Effortless Installation
- Ease of Management
- Peace of Mind



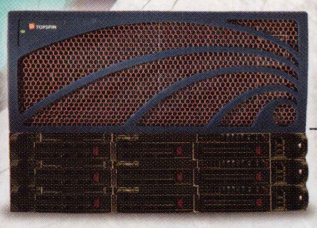
Servers

- 1U, 2U, 4U Options
- 32 and 64-Bit Platforms
- Cost Effective



Storage

- 36 Hot-Swap Drives in a 4U
- Serial ATA Technology
- Scales to Your Needs



Database Solutions

- Topspin Switched Computing
- Oracle RAC 9i and 10g
- SAN Connectivity

Intel Inside logo, and Xeon is a registered trademark of Intel Corporation or its subsidiaries in the United States and other countries.
© 2003, Promicro. All rights reserved. Promicro is a registered trademark of Promicro Systems, Inc.

promicro

intel premier PROVIDER 2003

Call us today at (866) 656-4276 or visit us on the web at www.promicro.com

is the box you build or buy will be faster.

The 8500-4 has four SATA-150 ports on the back edge of the card (toward where the drives are located). This arrangement, combined with the much smaller form factor of SATA cables, allows for much better airflow behind the drives. And these drives need some pretty serious airflow; the four 36GB Western Digital Raptor WD360GD Serial ATA drives run at 10,000 RPM. We deliberately went with these, instead of larger but slower RPM drives, to optimize for speed over storage space. These drives also have jacks for standard Molex power connectors as well as the new SATA type adapters, which help the transition between platforms. One drive is mounted below the floppy; the other three are mounted vertically in the housing at the bottom of the case, in front of the fans.

Now, we arrive at the case itself. All of the previously named components, plus the ENERMAX 465P-VE-24P 460-watt power supply, are contained in a custom Lian-Li PC-6270 quiet case. We chose this case because it is capable of accommodating extended ATX motherboards. As usual, this case comes with all the Lian-Li bells and whistles, such as thumbscrew assembly almost everywhere. In addition, four 5.25" and three 3.5" bays are accessible from behind a removable door on the front; an additional horizontal bay and five vertical-mount drive bays are inside. The case also has a slide-out motherboard tray and a filtered air intake under the chin of the case for the twin hard drive fans. A second matched pair of fans pulls heat out of the back of the case. An adapter converts the lower five vertical-mount bays to a four-bay horizontal-mount arrangement, and a handy cable clamp mounted with adhesive foam on the bottom of the case keeps all the SATA cables out of the airflow. Two USB jacks are provided on the front of the case and are accessible when the door is closed. The case is kept quiet by the use of dense foam batting inside the top and side panels, and a rubber seal around the door also helps cut drive noise.

Now that we have all the hardware bases covered, we need an OS. I chose SuSE Linux Enterprise Server 8 (SLES 8) based on what I had seen in SuSE's 32-bit offerings. Although SLES 8 isn't a bells, whistles and eye-candy OS, I was impressed by its smoothness and ease of use. It comes with base GNOME and KDE environments and a full set of development tools, making it an excellent base on which to install or develop professional applications.

If that weren't enough, YaST2, SuSE's GUI setup tool, has completely sold me on SLES 8. YaST2 does everything from printer configuration and user additions to advanced firewall configuration, and it does so quickly, clearly and with few surprises. The only thing I found that it could not do was configure a non-ALSA sound card—more on that later.

Often, when installing many distributions on anything other than the x86 architecture, packages are out of date, things don't work quite right and other annoyances abound. SLES 8 comes with kernel 2.4.19, KDE 3.0, Samba 2.2 and XFree86 4.2. These packages aren't bleeding edge, but quite current. They work well together, too, and aside from that one sound annoyance, I didn't find anything wrong that I couldn't fix easily.

As I said, getting sound to work on this system was an adventure. The Audigy 2 is relatively new, and I had a feeling that getting it to work might be a chore. I was not disappointed. The ALSA configurator in YaST2 correctly identified the card, but no sound came out. I tried using the emu10k1 Project available on SourceForge. After hand-tweaking /etc/modules.conf for the latest version of the OSS driver, the module loaded, but the system still had no sound. Prior to giving up and submitting a bug, I searched the SourceForge bug database and found that the CVS version was reported to work. But, would it work on a 64-bit platform? After a few make scripts, I saw the CVS version conflict with the Realtek chip on the motherboard, so I took its listing out of modules.conf and rebooted. The subsequent strains of Theodorakis' "Ode to Zeus" pouring forth from the speakers heralded victory.

Lack of time and equipment precluded any sort of high-end technical test, but I consider myself a decent audiophile, so I set up a test to satisfy myself that the Audigy 2

NVIDIA: PROPRIETARY BUT RESPONSIVE

While putting the finishing touches on this article, I received a call from Jeff Brown, of NVIDIA, following up on getting their driver to behave with the Ultimate Linux Box. I asked him to comment on why NVIDIA had not open-sourced their driver, which might have enabled a faster resolution of my issues. His response was the driver—the core of which is a single code base that works on everything from Windows to Linux to FreeBSD to OS X—was about 50% of NVIDIA's "intellectual property" content, the rest of the IP being the GPU itself. By comparison, on network cards, Intel appears to have most of the smarts on the card itself and subsequently has GPLed the driver. In compromise to the Open Source community, NVIDIA has committed itself to keeping support and feedback channels open. Brown said that Andy Mecham, one of NVIDIA's engineers, devotes half his workday to providing help to folks on the Linux forum, nvnews.com. I saw evidence of this myself in my search for answers; Andy's name seemed to be fairly common on the boards. Brown also said the folks that really pay the freight on the Linux side, members of the Visual Effects Society (of the members he mentioned, the name Disney stuck in my head) are satisfied with NVIDIA's support. These are the people that would use a commercially supported system similar to the Ultimate Linux Box 2003 as a multimedia workstation to do bleeding-edge stuff.

The fact that Monarch was able to get me, a solo end user, direct access to an engineer who did pin down the problem correctly—without having access to my machine—speaks well of NVIDIA's commitment to support. Although I don't think we're going to see a GPL driver for any NVIDIA card anytime soon, I think perhaps we're getting the next best thing—a company that knows where the future is and is committed to helping us get there without giving away its secret recipe.

is worth the investment. My wife and I have side-by-side machines with identical speakersets. I put the sound cable from her system in the appropriate jack on the Ultimate Linux Box and started Arnaud's "Bulger's Dream" from CD. I used XMMS with the EQ off for both tests, and I have to say, qualitatively, the Audigy 2 beat my Esoniq 5880 hands down for clarity and frequency response. It was quite a bit crisper and didn't develop any distortion or hum at high volume settings. Cards are available that probably can do quite a bit more than the Audigy 2, but we know this one is a good starting point. Plus, it works on the 64-bit platform, which is an achievement in itself.

Prior to press time, we were able to file down one rough edge—the video card. If you read my August 2003 Web article [www.linuxjournal.com/article/6922] on the subject, you may recall we rejected ATI's Fire GL X1 because it has drivers that work only on 32-bit platforms. My source at ATI was unwilling to comment on record about a 64-bit release. The ATI representative did at least ask me, unsolicited, what card I was running and why I had switched cards after reading my Web article.

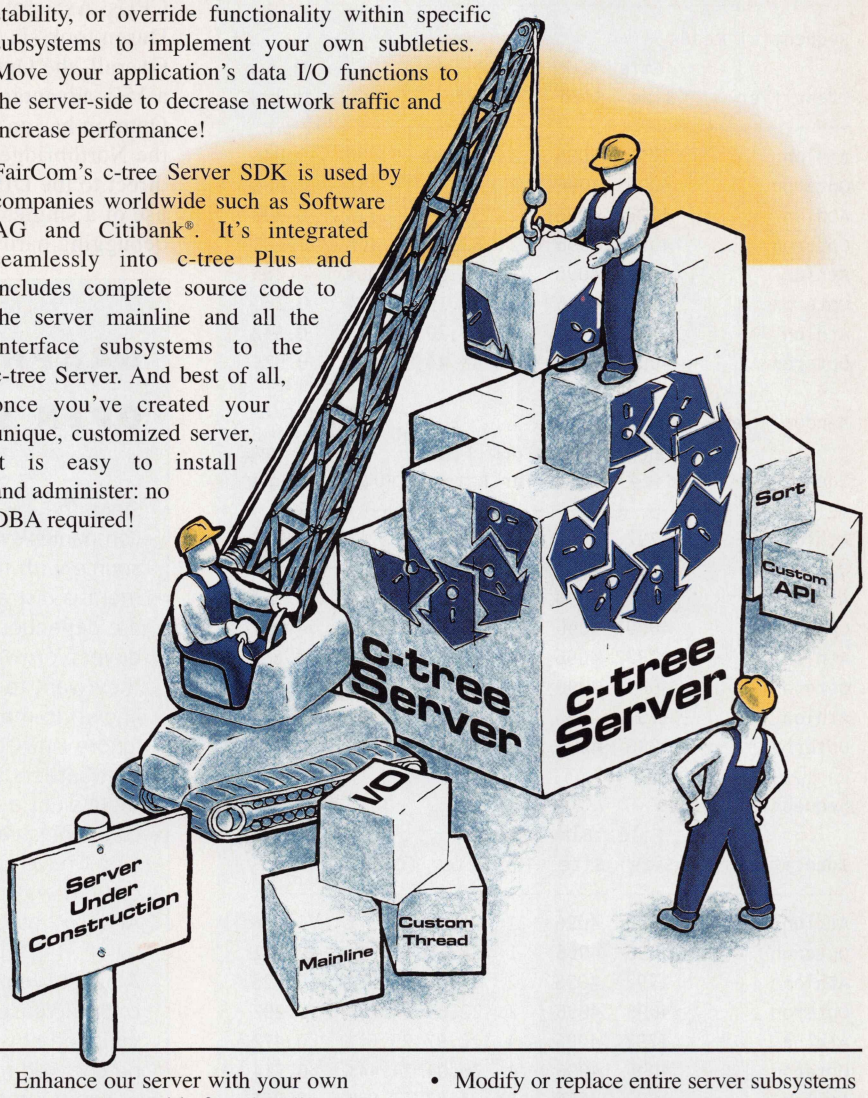
When I initially tried to run the driver for the NVIDIA Quadro FX 1000 card, the X server hung in max CPU mode, even though the driver had been available more than six months. Monarch put me in touch with NVIDIA directly. The NVIDIA engineer, Mark Visconti, took a look at my current information and suggested I upgrade the BIOS, which appeared to be an engineering sample version. I dutifully downloaded the BIOS and flash utility, both of which refused to work. Fortunately, from previous issues with the motherboard, I had a contact at Arima, the board maker. As it turns out, some arcane arguments have to be given to PHLASH.EXE—even now, you still have to boot DOS to flash the BIOS—to get the new image to load. After this step, I went back and reset the BIOS settings Visconti recommended. When booting into Linux this time, `startx` finally rewarded me with the cautions-bomb-boxes background that is a root X login.

CUSTOMIZE
YOUR DATABASE SERVER WITH THE C-TREE® SERVER SDK

Today's database demands are often too complex for traditional database servers. The functionality and precise level of control you need is simply not available. Perhaps you need alternate sort criteria for your data or a special twist in the threading or communication logic.

FairCom's c-tree® Server SDK allows you to create a customized, industrial-strength server designed for your particular needs. Use FairCom's kernel, with over 20 years of proven stability, or override functionality within specific subsystems to implement your own subtleties. Move your application's data I/O functions to the server-side to decrease network traffic and increase performance!

FairCom's c-tree Server SDK is used by companies worldwide such as Software AG and Citibank®. It's integrated seamlessly into c-tree Plus and includes complete source code to the server mainline and all the interface subsystems to the c-tree Server. And best of all, once you've created your unique, customized server, it is easy to install and administer: no DBA required!



- Enhance our server with your own custom server-side functionality
- Move functionality from the client-side to the server-side to reduce network traffic and increase performance
- Modify or replace entire server subsystems
- Complete source for the server mainline, key server subsystems, and client-side
- Flexible OEM licensing

Visit www.faircom.com/ep/lj/sdk today to take control of your server!

FairCom®
www.faircom.com

USA	573.445.6833
EUROPE	+39.035.773.464
JAPAN	+81.59.229.7504
BRAZIL	+55.11.3872.9802

DBMS Since 1979 • 800.234.8180 • info@faircom.com

Other company and product names are registered trademarks or trademarks of their respective owners. © 2002 FairCom Corporation

I was able to verify that the server was running in 3-D mode, but we did not have time before going to press to do video benchmarks. Given our tests of the NVIDIA card on a 32-bit machine, we should see frame rates in the mid- to high 50s, if not higher. If we do manage to get it tested at some point, you'll be able to find those results on our Web site.

The rough edges on this machine, however, are only in the eye candy. Aside from Chromium (the video test), we were able to run the rest of the *Linux Journal* Benchmark Bundle on the machine (bonnie++, postgres-test, tiobench and the kernel compile), and the results were impressive. The base kernel compile averaged one minute 35 seconds using both CPUs. As a point of comparison, my personal 1.1GHz Duron took six minutes 50 seconds to do the compile. Overall, the Opteron/3ware architecture seems to offer about a 15% advantage in speed, gigahertz for gigahertz. The Opteron has an integrated memory controller that bypasses the Northbridge and gives you a 64-bit duplex channel direct to the DIMMs. This particular machine also allows use of a singleton DIMM in 32-bit mode, which is handy for debugging hardware problems, at the expense of speed.

Listing 1. Tiobench Output (edited for space)

Sequential Reads

Identifier	File Size	Blk Size	Num Thr	Rate	(CPU%)	Avg Latency
Athlon	1792	4096	1	81.60	29.29%	0.048
Opteron	4096	4096	1	59.30	14.86%	0.066
Athlon	1792	4096	2	58.07	30.61%	0.132
Opteron	4096	4096	2	62.18	13.46%	0.125
Athlon	1792	4096	4	54.37	61.00%	0.285
Opteron	4096	4096	4	59.19	14.59%	0.260
Athlon	1792	4096	8	55.29	64.72%	0.542
Opteron	4096	4096	8	48.44	13.17%	0.625

Random Reads

Identifier	File Size	Blk Size	Num Thr	Rate	(CPU%)	Avg Latency
Athlon	1792	4096	1	1.32	0.975%	2.952
Opteron	4096	4096	1	1.18	0.151%	3.296
Athlon	1792	4096	2	2.29	1.374%	3.354
Opteron	4096	4096	2	1.93	0.740%	4.017
Athlon	1792	4096	4	3.18	2.859%	4.639
Opteron	4096	4096	4	2.76	1.236%	5.431
Athlon	1792	4096	8	3.70	2.221%	7.264
Opteron	4096	4096	8	2.96	2.085%	9.860

Sequential Writes

Identifier	File Size	Blk Size	Num Thr	Rate	(CPU%)	Avg Latency
Athlon	1792	4096	1	20.65	11.17%	0.126
Opteron	4096	4096	1	28.15	10.78%	0.101
Athlon	1792	4096	2	22.15	26.81%	0.228
Opteron	4096	4096	2	22.69	14.23%	0.292
Athlon	1792	4096	4	22.97	29.67%	0.472
Opteron	4096	4096	4	20.04	14.44%	0.714
Athlon	1792	4096	8	21.87	27.93%	0.856
Opteron	4096	4096	8	13.42	11.03%	1.978

Random Writes

Identifier	File Size	Blk Size	Num Thr	Rate	(CPU%)	Avg Latency
Athlon	1792	4096	1	0.60	0.234%	0.014
Opteron	4096	4096	1	0.47	0.121%	0.009
Athlon	1792	4096	2	0.59	0.479%	0.028
Opteron	4096	4096	2	0.50	0.159%	0.011
Athlon	1792	4096	4	0.64	0.542%	0.029
Opteron	4096	4096	4	0.49	0.155%	0.012
Athlon	1792	4096	8	0.68	0.558%	0.036
Opteron	4096	4096	8	0.50	0.192%	0.013

GRAPHICS DRIVERS AND OPEN SOURCE

Since the absorption of 3dfx, all three major graphics companies—ATI, NVIDIA and Matrox—have gone closed-source with their 3-D accelerated drivers. This development is distressing for several reasons. First, we now are dependent on the vendors to develop—and fix—the drivers, on whatever schedule they wish to follow. If they want to release drivers first for every other OS known to man and make us wait two years, or even ignore Linux entirely, that's their choice. Given the current barriers to market entry in the graphics field, there's not a lot you and I can do about it. Even if these companies choose to support us on a timely basis, we won't have access to a lot of things, including beta code. It was only by pulling CVS code that I made the Audigy 2 play.

A purist would point out that not only does releasing code increase the number of people that can debug your code by orders of magnitude, it also means it can be checked for funny business, which has been a topic of some interest in graphics drivers of late. Finally, there isn't much of a reason not to release the code; it's not like you can use it without the card. I don't know the official reasons behind keeping the drivers proprietary—I've been too busy trying to get them to work to find out. But I would like to see the reasons debated openly. I think the Linux community and the graphics people can negotiate an amicable, open settlement whereby we all can get what we need: the best drivers on the best cards on the best operating system. I suspect the first company to come to the table will see an improvement in its bottom line as well, because Linux folk tend to back up their principles with their hardware-buying budgets.

Tiobench produced the other interesting numbers as shown in Listing 1. I compared the performance of the 3ware/Western Digital harness on an Athlon 2800 to its performance on the ULB. The numbers are not quite apples to apples, because I took advantage of the Opteron's 64-bit addressing to handle a larger file size—but that in itself produced interesting results. The 32-bit platform appeared to do better in single-threaded sequential reads but produced comparable numbers for the rest of the run. On random reads and writes, the absolute rate of the ULB lags behind, but the latency does not. The ULB soundly trounces the Athlon despite using a file more than double the size. I suspect that the drop in numbers at the end of the sequential read and write runs is an artifact of hitting the 3ware card's buffer limit, given the nice flat curve up to that point. On a side note, these numbers compare favorably in most departments in absolute terms—and very well in terms of bang for buck—with a Dell Precision 650n SCSI system I tested previously.

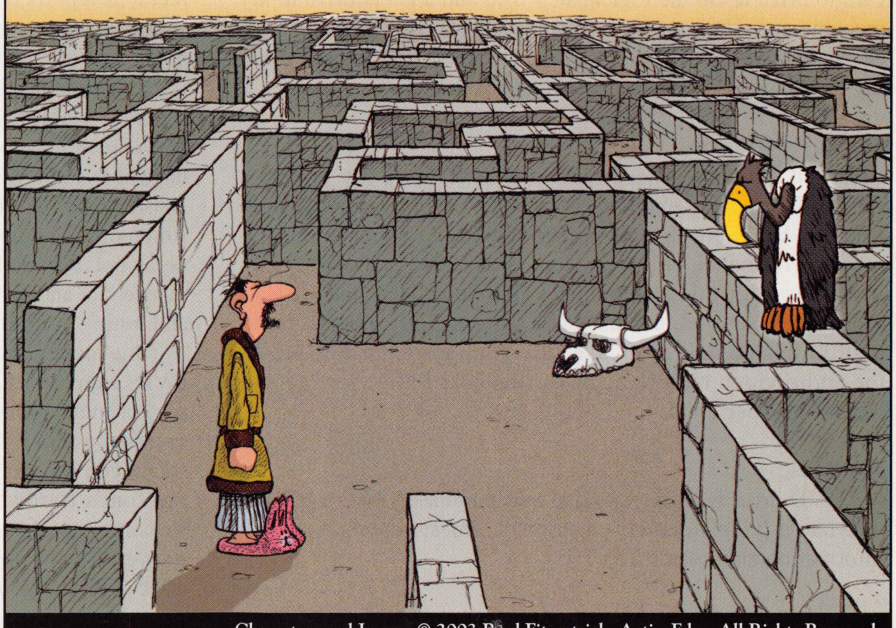
After receiving some feedback on our Web article series that discussed the ULB, we decided we needed to test noise levels. The ULB scored 50.5dBa at 10" in front of the case, 50.0dBa at the operator position (24" above the top of the case) and 60.0dBa at 10" from the back of the case. Obviously, these numbers aren't as low as we'd like them to be with the aforementioned Dell coming out at 47, 45 and 55dBa respectively. I think the culprit is the Thermaltake coolers; when we had the machine in-house previously with AMD coolers, it was quieter by far. However, the AMD version did get rather warm. The ULB version seems to be quite stable with respect to temperature, at the expense of extra noise. I suspect that between now and when this issue is on newsstands, companies such as Zalman and PC Power and Cooling will have offerings available for the Opteron like the ones they now have for noisy Athlons.

The Ultimate Linux Box, like Linux itself, is a work-in-progress. By the time you read this, new toys will be on the market, not to mention new software, maybe even a major kernel revision. The Escalade 8506 also will be available, as will newer, bigger Western Digital Raptors, DVD-plus-or-minus-RW combo drives—perhaps even an Athlon 64 CPU and motherboard. Use this design as a place to start and make your own improvements.

The author would like to thank 3ware, Western Digital, NVIDIA, Arima and SuSE for their contributions and Monarch Computer Systems for tying all the hardware together in one package and making it play.

Glenn Stone is a Red Hat Certified Engineer, sysadmin, technical writer, cover model and general Linux flunkie. He has been hand-building computers for fun and profit since 1999, and he is a happy denizen of the Pacific Northwest.

CODY HAD PASSED THE SKULL MANY TIMES, BUT THE BIRD WAS NEW



Characters and Images © 2003 Brad Fitzpatrick, ActiveEdge. All Rights Reserved.

Etnus TotalView... *The World's Best C/C++ Debugger*

Debugging C++ programs can leave you feeling a bit lost and disoriented. But TotalView shows you hierarchical class derivation, gives you superior support for templates, handles shared libraries and DLLs, lets you test fixes on the fly and much more. So don't get lost in a C/C++ labyrinth.

Try TotalView FREE at www.etnus.com

ETNUS TOTALVIEW

DVD Players

We compare features, convenience and performance of the leading DVD-playing software for Linux and cover some important tweaks for smooth playback. BY DAVE PHILLIPS

I love movies. I'm a fan of films from almost every genre, from various countries and in various languages. Last Christmas, Ivy and I received a DVD player from her kids, and since then I've become an unabashed convert to the medium. Images are sharper, sound is clearer and the medium itself permits my choice of amenities such as subtitling, scene selection (aka chapters) and language preference. The only real problem I have with the player is it resides at her house and not mine. I don't even own a television, so purchasing a standalone DVD player also would mean buying a TV, and that's a purchase I'd rather not make. However, I do have a nice 19" monitor attached to my computer and a good video card to drive it. So the logical step is to add a DVD drive to my machine and configure my system for DVD playback. This article describes how I did that, the problems I encountered and my impressions of the software used for the job.

The Test Platform

I learned a lot about hardware while setting up my system for DVD play. My machine is capable of enjoyable glitch-free playback, but it took some tweaking to squeeze the most performance out of it. Here, I describe my system as it was used for this article, but I also describe what I recommend for a more current base system. Hardware matters a lot in this domain, so make sure your system can handle the audio/video requirements for the best DVD viewing and listening experience.

The Real

My DVD drive is an inexpensive unit purchased from a friend who had it lying around his apartment. dmesg reports this information about the drive:

```
hdd: LITEON DVD-ROM LTD163D, ATAPI CD/DVD-ROM drive
```

If you're in doubt as to what kind of drive is in your machine, run `dmesg | grep DVD` for a report. Linux has exceptional support for DVD drives; in fact, the Linux Hardware Compatibility HOWTO states that virtually all ATAPI and SCSI DVD-ROM and DVD-RW drives are supported. If you're looking for recommended brands or are concerned about a specific drive, a quick search on Google should turn up the needed information.

The base machine also includes an 800MHz AMD Duron processor, two 15GB hard disks and 512MB of RAM. The display hardware is a generic 19" monitor connected to an NVIDIA GeForce2 video card with 64MB of video RAM. Audio is handled by a Sound Blaster Live! Value card and a sound system that includes a Yamaha DMP7 digital mixer, a 100-watt QSC power amplifier and a pair of Yorkville Sound YS-10 studio monitors. The kernel is compiled for low latency, and all drives are tuned for optimal disk throughput.

The Recommended

I prefer a faster CPU, something 1GHz or higher. I have received reports of decent DVD playback on 600MHz and slower machines, but at less than 1GHz you need to tune your other system components more finely. If you like to watch films in wide-screen mode, I recommend at least a 19" screen for comfortable viewing. Your graphics chipset should support the XFree86 Xv extension (most do), and your card should have at least 16MB of video RAM. The Creative Labs Sound Blaster Live! or Audigy 2 sound cards are excellent choices for stereo or 5.1 audio output configurations. The low-latency kernel is not absolutely necessary, but your DVD drive performance can and should be optimized. I discuss kernel options and tuning your drive in the next section.

One more note regarding an optimal audio system. Many DVDs support 5.1 Surround Sound and other audio options that may or may not be possible under Linux. An inexpensive 5.1 speaker system is fine for casual use, but if you're serious about sound you probably want to invest in a high-quality system. See the Tom's Hardware Guide URL in this article's Resources section on the *Linux Journal* Web site (www.linuxjournal.com/article/7174.html) for more information regarding available 4.1 and 5.1 sound systems for PC sound cards.

Configuration Considerations

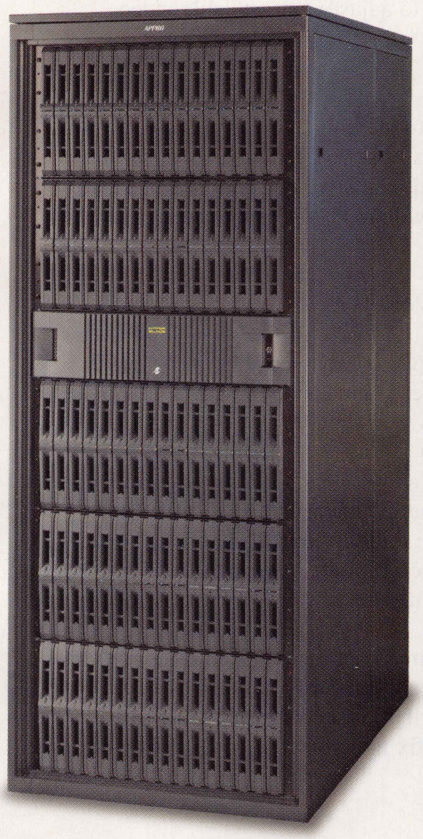
The base Linux installation used for these tests is a heavily modified Red Hat 7.2, but the methods and procedures described here should apply to any relatively recent mainstream distribution with only minor changes (if any). Plain-vanilla Linux is not likely to yield optimal results for DVD performance, so following are a few tips I gathered from the Web and from the documentation for various players tested. The main considerations concern optimizing the kernel itself,

Appro High-Performance Servers & Cluster Solutions

Appro's servers based on AMD Opteron™ processors boost performance on existing Windows® & Linux applications.

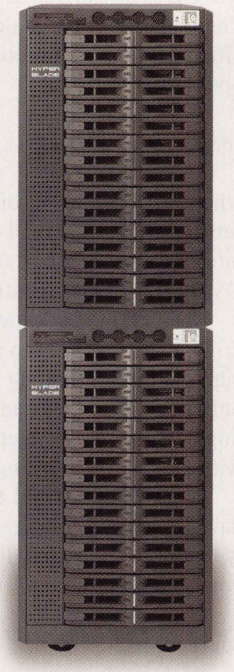
Double the Density!

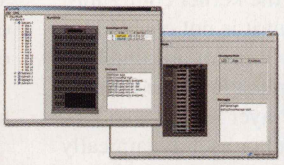
Appro *Hyper*Blade Fully Integrated Cluster Solutions



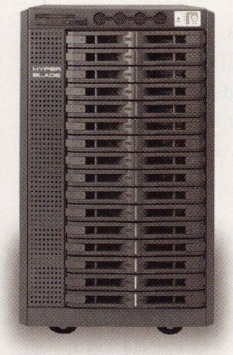
New!

Mini-Cluster



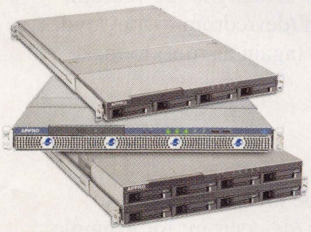


Appro BladeDome Remote Server Management Option



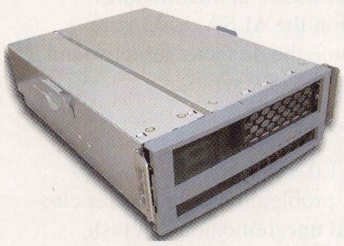
- Flexible, modular and scalable architecture for rapidly changing environment
- Blade servers fit Appro full-size cabinet (up to 80-nodes) and Appro mini-size cabinet (up to 17-nodes) with the ability to mix and match cluster building blocks
- Dual AMD Opteron™ processors per node
- Up to 16GB ECC DDR Memory per node
- One PCI-X slot per node
- Up to two internal HDDs
- Remote Management Options:
  - In-band and out-of-band control
  - GUI & Command Line Interfaces
- High-Speed Interconnect Options: Myrinet, Dolphin, Quadrics and Infiniband

Appro 1U and 2U Servers



- Dual AMD Opteron™ processors
- Design to run cool and efficiently in demanding rack conditions
- High quality, powerful and cost-effective
- Fully integrated cluster solutions available

Appro 4U - 4 Way Server



- Quad AMD Opteron™ processors
- Memory capacity and flexibility to support a wide range of applications
- Supports up to 64GB in 16 DIMM sockets
- Supports up to 4 SCSI drives
- 3.5" floppy drive and slim CD-ROM

AMD Opteron™ Processors - Shorten run-time cycles and increase bandwidth for processing computing requests.
- 32 bit applications while you migrate to 64 bit computing for long-term investment protection.

APPRO
HPC Cluster Solutions

For more information, please visit **www.appro.com**
or call Appro Sales at **800.927.5464, 408.941.8100**.

AMD 64 Opteron

tuning X, tuning the DVD drive and ensuring you have the correct device mountpoint.

Kernel Options

Some kernel options should be activated for optimizing DVD playback. I strongly urge the application of Andrew Morton's low-latency patches, and it also may be advisable to apply Robert Love's preemptive kernel patch. The combination of these patches provides very low latency (below 3msecs) over sustained periods of time. My kernel of choice is currently 2.4.18, but Andrew's and Robert's patches are available for a variety of kernel releases (see Resources for more information).

You also should make sure your kernel has enabled support for the RTC and MTRR options (found under the Character devices and Processor type and features kernel configuration sections, respectively). RTC provides access to the real-time hardware clock of your PC. According to the kernel configuration help file, all PCs have such a clock but it is not enabled by the default kernel configuration. Although it is not absolutely required, many Linux audio and video applications can utilize this clock for a finer timing response (MPlayer likes it), so I suggest building it directly into your kernel or as a dynamically loadable module.

According to the kernel configuration help, enabling the MTRR (memory type range registers) provides a mechanism that is used:

> ...to control processor access to memory ranges. This is most useful if you have a video (VGA) card on a PCI or AGP bus. Enabling write-combining allows bus write transfers to be combined into a larger transfer before bursting over the PCI/AGP bus. This can increase performance of image write operations 2.5 times or more.

So, if you have a PCI or AGP video card you will want to enable this option.

Under the ATA/IDE/MFM/RLL Support section, I advise enabling the options for generic PCI bus-master DMA support and the use of PCI DMA by default.

The last step is to configure sound support for your hardware. I use the ALSA sound system, so all I do is enable sound card support in the kernel options. The ALSA drivers are built in normal user space and installed as root as loadable kernel modules. You safely can use the available kernel modules instead of ALSA, but in my opinion the ALSA drivers are superior. In fact, ALSA will become the de facto kernel sound system for Linux kernels beyond the 2.5.x series.

The X Factor

I began my tests with XFree86 4.1.0. Everything seemed to work fine except for an annoying problem with xine: after closing that player my X server would unceremoniously crash, dumping me back at the console prompt. When I upgraded to XFree86 4.3.0 and the latest driver for my GeForce2 from NVIDIA (1.0-4363), all problems were resolved. Because XFree86 4.3.0 fixes a number of problems and bugs found in the earlier versions, I suggest the upgrade to anyone using any of these players, not only xine. And, if you're using an NVIDIA card you always should use their latest drivers.

As mentioned earlier, XFree86 provides a video output driver called Xv, but other drivers are available for the framebuffer device, SDL (Simple DirectMedia Layer) video output and X11/Xshm. The driver you select affects playback performance. Xv normally is preferred, but you can experiment with whatever drivers are supported by your player of choice for the best output.

My X environment also includes the Blackbox window manager. I prefer simple and fast, and Blackbox fits the bill for me. Be advised that your DVD-viewing mileage may vary in part due to your window manager or desktop environment of choice, and you may need to adjust your available video resources accordingly.

Tuning the Drive

It may come as a surprise to learn that you can tune your DVD and CD-ROM drives as easily as you can tune your hard disks. The hdparm utility can optimize drive performance to peak efficiency, run it (as root) with at least the following parameter options:

```
hdparm -c1 -d1 -a8 -u1 /dev/hdd
```

where -c1 enables 32-bit I/O, -d1 enables DMA access, -a8 sets the filesystem read-ahead value and -u1 sets the drive's interrupt-unmask flag. /dev/hdd should specify your particular DVD drive device location.

The parameters shown above work well with my DVD drive, but I urge you to read the hdparm manual page (`man hdparm`) before running the utility. Your DVD drive is a read-only device, so filesystem corruption is not an issue here. You might, however, inadvertently lower your drive's efficiency with non-optimal settings.

It has been brought to my attention that Red Hat 8.0 users have reported difficulties enabling DMA on their DVD drives. If you're running that distribution, add the following line to your /etc/modules.conf to fix the problem:

```
options ide-cd dma=1
```

About /dev/dvd

The players reviewed here all expect to find the default hardware mountpoint at /dev/dvd. Though they all also allow a user-specified location, I suggest making things easier by making /dev/dvd. Typically the drive itself actually is /dev/cdrom, so you may need to create a link from /dev/cdrom to /dev/dvd. Simply issue the following command (again as root) to make the link:

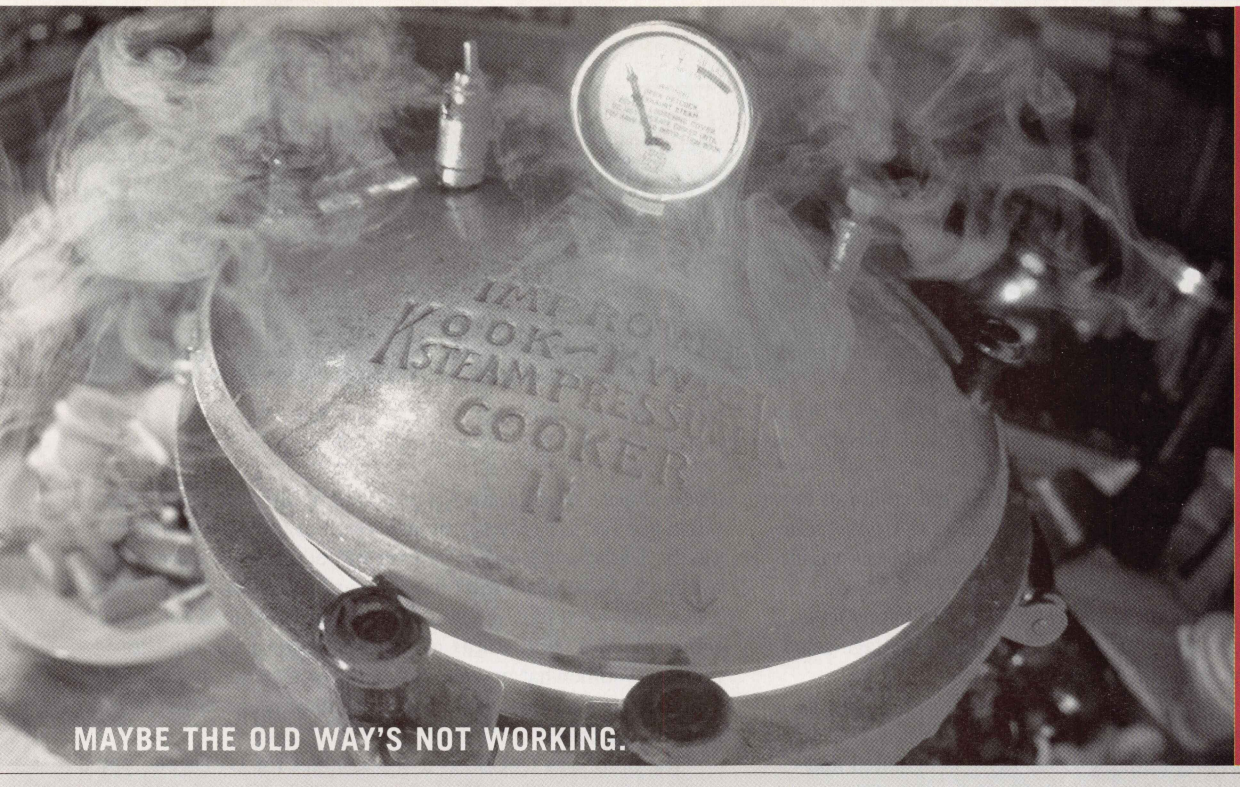
```
ln -sf /dev/cdrom /dev/dvd
```

If you have multiple CD/DVD-type drives, you need to specify the correct device number for /dev/cdrom; for example, mine is /dev/cdrom1 because my CD-RW drive sits at /dev/cdrom.

The Test DVDs

I tested the players with a variety of DVDs, all legitimately manufactured and purchased. My local library lends DVDs, many of which are in less-than-optimal condition, and they

DATA CENTER ABOUT READY TO BLOW?



Rackable Systems servers—now with Distributed DC Power—improve thermal management, reliability and power efficiency.

GREATER RELIABILITY & MINIMIZED MAINTENANCE

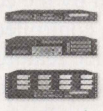
Now fueled by Distributed DC Power, Rackable Systems server clusters are more reliable, flexible and efficient than the competition. By distributing redundant DC Power to each server—and replacing the standard AC power supply with a 70 times more reliable DC card—availability is significantly improved and servers are easier to maintain. They also run cooler, harnessing the full power of the fastest Intel® Xeon™ processors. Best of all, Rackable Systems DC clusters accept either AC or DC input power for easy, immediate deployment. All at a price comparable to standard AC-based solutions.

THERE IS A BETTER WAY

As pioneers of large-scale data center solutions, Rackable Systems continues to lead the industry with innovative server designs for Biogen, Electronic Arts, Google, Inktomi, Lawrence Livermore National Laboratory, nVidia, SLAC, Tellme Networks, Webex and Yahoo!. Our open architecture configurations are built-to-order to ensure maximum flexibility and ease of serviceability. Our innovative cabinet design features twice the density at a fraction of the cost, while optimizing thermal efficiency. And with Rackable Systems' new Distributed DC Power technology, you've got a solution that can't be beat.

Rackable® systems

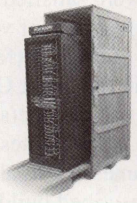
Flexible Configurations



- Customized & built-to-order
- Open architecture
- Superior remote management

Improved Reliability

- Redundant power & drives available
- Thermally efficient & optimized
- Reduced data center costs



Easy Deployment

- Fully configured & tested
- Pre-cabled, pre-racked
- Quick ship

Learn more at **www.rackable.com/dcpower** or call 1.866.548.2527.

intel inside XEON

IN WITH THE NEW

Rackable® systems

©2003 Rackable Systems. All rights reserved. Intel, Intel Xeon, Intel Inside, and the Intel Inside logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Table 1. Comparison of Linux DVD Players

Player	Version Tested	Subtitles	Menu Support	Random Seek	Keyboard Control	GUI*	CPU Usage**	License
MPlayer	0.90	Yes	No	Yes	Yes	Optional	40%–50%	GPL
Ogle	0.9.1	Yes	Yes	Yes	Yes	Optional	20%–40%	GPL
VideoLAN Client	0.5.3	Yes	Yes	Yes	Yes	Default	20%–30%	GPL
xine	1-beta12 (lib) 0.9.21 (ui)	Yes	Yes	Yes	Yes	Default	20%–40%	GPL

*Indicates whether a GUI is an optional or default feature of the build process. **The figures shown represent average low- to peak-CPU usage reported by gkrellm during play of the *Blade Runner* DVD. System load included XMMS, five active workspaces (under the Blackbox window manager) and an active DSL network connection running either Netscape or Opera.

played fine on the test system, with the single exception of an incredibly bad duplication of Bruce Lee's *Chinese Connection*. I even was able to watch a truly awful DVD of dubious origin, a remarkable event given that our standalone DVD player wouldn't even recognize the disc in its drive. I'm happy to report that in all tests the options for subtitling, language selection, chapter jumps and skins worked.

Linux DVD Players

So your kernel is configured, the DVD drive is installed and connected, and you're ready to watch *Shrek* for the 40th time. All you need now is a player application, and happily Linux has some excellent DVD player software. The profiles below focus on four of the most popular players: MPlayer, Ogle, VideoLAN Client and xine. See the Freshmeat listings for other available players and DVD amenities.

Expectations

I've already used a standalone DVD player, so I expect to find most of its features in whatever software player I select. In the players reviewed here, I looked for support for these minimum features: standard transport controls (start, stop, pause, fast-forward and rewind), scene selection, subtitling and audio preferences and DVD menus. In addition to the amenities found on the standalone hardware player, I expect a software DVD player to switch from windowed to full-screen view easily and to offer random seek/relocate, keyboard control of all transport functions and skin support.

Most of my expectations were satisfied by the players I reviewed. See Table 1 for an overview of versions, features, licensing and CPU stress. All of them performed with excellent results, with no clear winner in the "Best Of" category. My advice is to try them all, then use the one(s) that seem best to you. In terms of weight, Ogle is the lightest (it's a DVD-only player) while the others all come in with about the same metric tonnage. Although I briefly described building the programs, the reader should check the player Web sites for available RPMs and other prepackaged binaries.

Win32 Codecs and the DeCSS

Much of the magic performed by these players comes from their use of video codecs (compression/decompression libraries) found in Windows and the Mac OS. In particular, MPlayer and xine require certain codecs to play files in formats such as Microsoft's ASF/WMV and Apple's QuickTime MOV. These codecs usually are not provided by the source or binary packages for the players themselves, but they are acquired easily. I advise getting whatever collections are currently available. Most of us want the standard packages, but if you're a die-hard video fanatic you might as well download and install them all. The legal status of acquiring and utilizing these codecs is somewhat unclear, but because they are available now I suggest getting them right away. It might prove difficult to do so at a later date.

The notorious DeCSS is a descrambler for DVDs encrypted with the Content Scrambling System (CSS). The original DeCSS software was proprietary and binary-only, but it has been reverse engineered and has spread through the Internet. A great deal of ink and ill will has been spilled over CSS, and the DVD legal battles are far from over. For more information regarding the issues involved, please read the material at cyber.law.harvard.edu/openlaw/DVD. None of the players discussed here require DeCSS. The open-source libdvdcss is used for runtime CSS decryption and does not require a region-locked DVD player.

MPlayer

MPlayer is undoubtedly the most full-featured player reviewed here. It is considerably more than a DVD player, handling files in far too many video and audio formats to list here. Suffice it to say that if it's video, MPlayer likely can play it.

MPlayer's DVD support meets the standard expectations listed above, and it adds a few features I now can't live without, including its equalizer, a control for audio equalization (EQ) as well as video EQ. Brightness, contrast, hue and saturation can be controlled by the equalizer's sliders. Both audio and video EQ can be controlled in real time with smooth responses. Another nice feature is the ability to set the aspect ratio, that is, how much of your available screen space is occupied by the picture. I don't use this feature so much with DVDs, as most discs are available in wide-screen or standard formats, often on the same disc. Finally, MPlayer provides preset normal, double and full-screen modes for your viewing pleasure.

I really like MPlayer, but I must advise that its build procedure can be somewhat complex. Look over the myriad of configuration options carefully before building the program (see the results from `./configure --help`). For instance, the MPlayer GUI is not included by default and must be enabled explicitly. DVD menus are supported with libdvdnav, but the documentation for MPlayer

Figure 1. MPlayer with Blue Skin

0.90 indicates that menu support currently is not working. MPlayer's extensive documentation explains every aspect of the program, including compilation details. Look there before you post a complaint to the MPlayer mail list.

One more note concerning MPlayer: its developers are not fond of certain versions of GCC 2.96, nor do they particularly recommend NVIDIA graphics cards. Their stated position is to not answer questions from users with systems owning those components, which is a bit of a problem for me because I have an NVIDIA card and a version of GCC 2.96 on my machine. Nevertheless, I have compiled and used MPlayer under those build conditions and am quite happy with the results. If you have problems when compiling MPlayer with those factors, a Google search should resolve them. In all fairness, I must add that the MPlayer developer and user community is otherwise quite helpful.

Ogle

Unlike the other players reviewed, Ogle is strictly a DVD player—but what a DVD player it is. Ogle was the first player to support DVD menus, and its other amenities include bookmarks, time skipping, multichannel audio, SPDIF audio output (a digital audio format) and crop and zoom video. The bookmarks function is unique and sweet: I can stop anywhere within a movie, bookmark my position, then return to that position later simply by clicking on the mark. It may not seem like an exciting feature, but it is handy. With the Goggles GUI, Ogle also supports the option to start play upon opening; that is, Ogle automatically starts playing the disc in the drive.

This Season Get Cozy with Hurricane Electric Internet Services!

Full 100 Mbps Port
With Full Cabinet
$2,500/month

Colocation Full Cabinet
Holds up to 42 1U servers
$400/month

Colocate Your Server
Up to 7U or 12 inches
256 kbps ~ 70GB
$200/month

Order Today!
email sales@he.net or call 510.580.4190

HURRICANE ELECTRIC INTERNET SERVICES

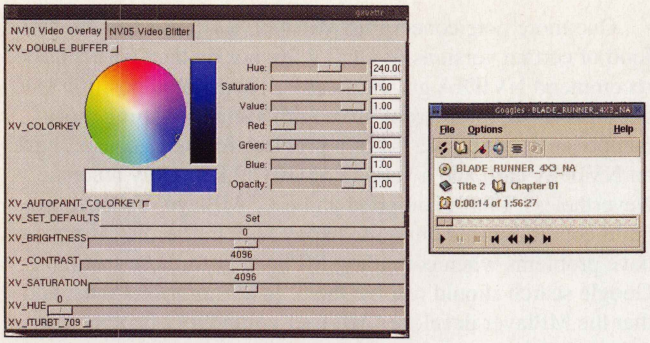


Figure 2. Ogle with Goggles GUI and gxvattr

Ogle by itself can be used from the command line. It also can be built with a native GUI, and a number of third-party GUIs are available from the Ogle Web site. Personally, I like having a control panel handy, and I especially like the appearance of the Goggles GUI, but Ogle's keyboard mapping is excellent. Also, the Goggles GUI requires the FOX toolkit, which is not commonly found in mainstream Linux distributions. The Ogle Web site can direct you to the Goggles home, where you can find out how to acquire the FOX toolkit.

I'm also happy that Ogle's documentation directed me to the xvattr utility. The xv driver typically is the default video output driver for Ogle and the other players reviewed here, and its capabilities usually can be modified by preference settings in the players themselves. However, xvattr is a standalone utility that queries your video card for its specific xv-related capabilities and allows direct user control over them from the command line (or the gxvattr GUI). I found it to be quite handy when trying to resolve some frame rate difficulties due to my NVIDIA card's default double buffering (I was able to switch it off using xvattr). I advise using xvattr to learn more about your video card's particular xv-related capabilities.

VideoLAN Client

VideoLAN Client (VLC) is one part of a project intended to provide a cross-platform client/server solution for A/V (audio/video) streaming over high-bandwidth networks. According to the excellent VideoLAN documentation, the project includes the VideoLAN Server (VLS), which can stream MPEG-1, MPEG-2 and MPEG-4 files, DVDs, digital satellite channels, digital terrestrial television channels and live videos on the network in unicast or multicast. It also includes the VideoLAN Client (VLC), which can be used as a server to stream MPEG-1, MPEG-2 and MPEG-4 files and DVDs on the network in unicast or multicast. It also can be used as a client to receive, decode and display MPEG streams under multiple operating systems.

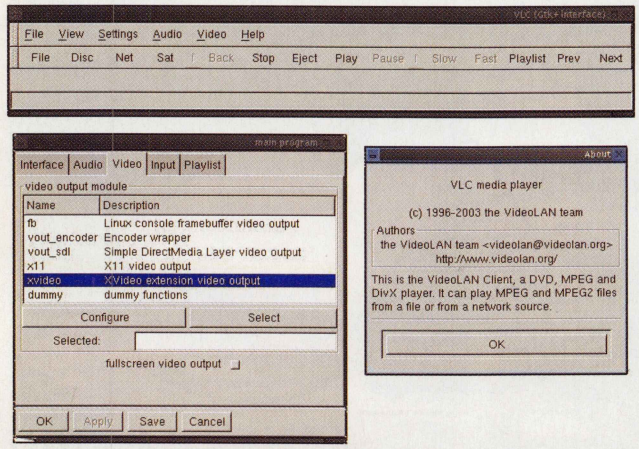


Figure 3. VideoLAN Client Controls and Preferences Dialog

As a standalone DVD player, VLC's performance is in line with the rest of the players reviewed. Its CPU usage was the lowest measured, making VLC a first-choice solution for networked machines or relatively low-powered systems. VLC's GUI isn't especially exciting, but it works smoothly and flawlessly. If you don't need extensive file format support or if your system fits the target model, then I have no hesitation recommending VLC.

xine

If MPlayer and xine were horses in a race, they'd be side by side at the finish line. xine is somewhat easier to build (the xine developers are not as parochial about your compiler and video card), and it includes all needed libraries within the source package, as well as its GUI in the default build. It also supports DVD menus by default, plays a wide variety of video formats and uses a video EQ similar to the one found in MPlayer. xine's GUI is a little strange at first, but it is actually well organized and easily navigated.

Figure 4. xine with default GUI

xine's performance is second to none. I was concerned about the results of my original test of its CPU usage, but I discovered the program remembered a "sticky" setting I had made while experimenting with its video output drivers. I had told xine to try the xshm driver, not realizing that subsequent sessions would continue to use that driver. Nothing was wrong with the driver itself, I simply noticed that xine's CPU usage was much higher than the results for the other players. Resetting the output driver to xv dramatically lowered CPU consumption, so I've been a little more careful with any changes I make to xine's default settings. You also can run the handy xine-check

for a résumé of your system's capabilities analyzed with regard to xine's performance.

The only feature I miss in xine is an interface for random chapter selection. I can jump to scenes before or after the current location, but at this time there is no graphic representation of the chapters tree à la Ogle or MPlayer. However, scene selection from the DVD menu itself is fully supported, so as long as the disc includes a menu with scene selection, xine performs random chapter jumps. One other item of possible complaint is xine's inclusion of its required libraries within the source tree. Apparently this bothers some people, but I consider it to be a great convenience. I didn't have to run around the Web to find what I needed to complete the build; all I needed (except the Win32 and Apple codecs) was included with the original package.

Incidentally, although xine's default GUI is handsome and perfectly usable, a variety of alternative user interfaces can be found on the xine home page. Some nice-looking GUIs are shown, including one that shows your video output in an ASCII character display.

As with the other players, xine's community maintains a number of highly active mail lists. If you have questions about xine that are not answered in its excellent documentation, you certainly can find help from its community of developers and users.

Ripping, Burning and Hardware Decoding

Rather than trying to rewrite the excellent and exhaustive advice found at bunkus.org, here is some brief advice regarding ripping a DVD. Get a big hard disk, install either mencode or transcode (and its dvd::rip GUI) and follow the detailed instructions on the bunkus site. Ripping a DVD can involve a large number of options, so plan ahead for best results. The author of the bunkus site recommends at least 10GB of free space per disc ripped. Also, even using a fast CPU, the ripping process can consume many hours.

I don't own a DVD burner, so I can give no useful advice regarding the process other than mentioning that Jörg Schilling's excellent **cdrecord** is at the heart of it. However, as with ripping DVDs, a number of on-line articles are listed in the Resources section on the Web (www.linuxjournal.com/article/7174.html) that describe the process in some detail.

While preparing material for this article I asked members of the Linux Audio Users mail list what DVD software they used. MPlayer and xine were the clear winners, but one respondent asked whether I intended to cover hardware DVD decoding boards. Alas, I have no experience with such hardware and welcome feedback from any readers who have used them.

Documentation and Support

Apparently none of the developers of the players reviewed ever wanted to be accused of supplying inadequate documentation. MPlayer, VideoLAN Client and xine supply especially extensive docs for developers and normal users. Ogle provides somewhat less exhaustive documentation, but it's also the most narrowly focused player reviewed, and its manual page (`man ogle`) is excellent. All of these players have extensive support available through highly active mail lists and list archives.

The Final Frame

The players reviewed all show remarkable longevity and maturity. Linux may not be an obvious first choice for a multimedia platform, but it's certainly becoming difficult to ignore the fact that it is rapidly evolving into a superb platform for audio and video play. I encourage readers to check out the software listed and reviewed here, and I look forward to receiving your reports. Have fun, but remember to try to get some real work done too.

Acknowledgements

My thanks to all the development teams working to bring a better DVD experience to Linux users everywhere. Thanks also to Siggi Langauf and Bill Fink for their assistance with xine and their astute criticism of my first drafts of this article. I am responsible for remaining inaccuracies and errors, and I welcome civil corrections and addenda.

Dave Phillips is a musician, teacher and writer living in Findlay, Ohio. He has been an active member of the Linux audio community since his first contact with Linux in 1995. He is the author of *The Book of Linux Music & Sound*, as well as numerous articles in *Linux Journal*.

BIGGER RACKS. LOWER T1 PRICING.

1U, 4U and 1/4 rack co-location services that can't be beat.

- OUR 1/4 RACKS ARE 15U (~26 INCHES)
- WE HAVE SOME OF THE LOWEST PRICES ON THE WEST COAST

1u to 2u	$50/month
4u	$75/month
1/4 rack	$150/month

ALL PRICES INCLUDE BANDWIDTH

LAYER 42

WWW.LAYER42.NET
INFO@LAYER42.NET
408.450.5740

DVD Authoring

Trick out home videos with a fun, featureful menu system that viewers can navigate from a regular DVD player. **BY IAN POINTER**

Traditionally, DVD authoring has been an expensive affair. Full-featured professional applications can cost thousands of dollars, while cheaper products, such as Apple's iDVD, have arbitrary restrictions that significantly reduce their usefulness. A new open-source effort, dvdauthor, is bringing the possibility of low-cost, professional-grade DVD authoring to Linux. Although it doesn't yet support all the features of the DVD specification, development is proceeding at a fast pace, and new features are being added with each release. Together with a more established open-source toolkit, mjpegtools, this article explains how to construct a relatively complex DVD application, a photo album, with dvdauthor. We also illustrate the various features that dvdauthor currently supports and how to use open-source tools to construct a DVD-R that can play on almost every DVD player.

How a DVD Works (Quick Version)

A DVD is comprised of one or more video title sets (VTSes), which contain video information in the form of MPEG-2 video streams. Each disc can have up to 99 VTSes, and each title set can be subdivided further into as many as 99 chapters, allowing DVD players to jump to a certain point within the video stream. Within each VTS, a DVD can have up to eight different audio tracks and 32 subtitle tracks that the viewer can switch between at will. A menu system can be included within a title set, allowing the viewer to select between the different subtitle and audio tracks. An optional top-level menu, known as the video manager menu (VMGM), is used to navigate between the different title sets. One VTS may contain a feature film and another may contain a documentary on the film, and the VMGM allows viewers to select which one they want to watch.

The DVD format doesn't eliminate the differences between the two competing broadcasting formats, NTSC (primarily used in America) and PAL (the standard in Europe and Japan). I live in Britain, so the frame information and resolution details used in this article are for a PAL system, but I point out the differences you need to be aware of when they appear and offer appropriate settings for an NTSC disc.

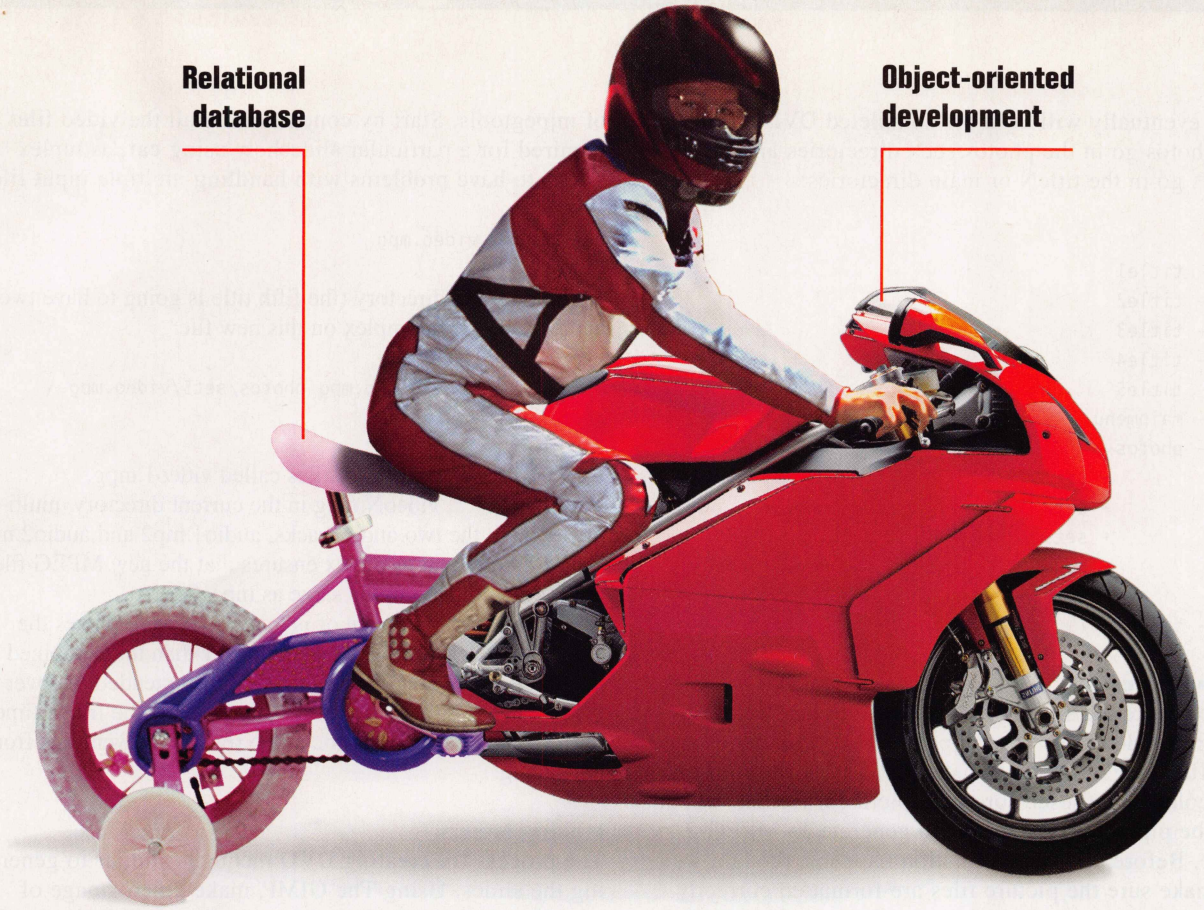
The DVD specification includes advanced features, such as the concept of region coding, the possibility of viewing different angles of a video stream and simple computations using built-in registers provided by a DVD player. I don't know much about these features, and they aren't discussed in this article. The dvdauthor mailing list is a good source for further information.

Planning

Before we rush headlong into creating menus, subtitling and multiplexing, it's a good idea to sketch out the structure of the DVD with paper and pencil. Proprietary DVD tools offer GUI systems for creating this type of structure, but no such tools are available yet for DVD production on Linux. As you'll soon see, the command-line tools have a lot of different options, so having your ideas on paper is preferable to trying to keep everything in your head.

The DVD application I'm creating is a photo album, using pictures that I took while studying abroad at UNC-Chapel Hill this past year. For simplicity's sake, I have only six photos in each category. On paper, I decide that the main menu (the VMGM unit) should have five buttons, four of which are simple text buttons (one for each different photo category), plus a secret link unlocking extra pictures (secret extra features are a common occurrence in commercial DVDs) and a music track playing in the background. The four regular buttons link to one of four menus, one for each different section. The menu system for each section consists of two menus and an audio track, with selectable preview images of the slideshow, a button to move onto the next set of preview images and two buttons that allow the viewer to watch the complete slideshow or go back to the main menu. To keep things simple, the photo slideshow should have the same song as the section menu playing in the background. After the slideshow is finished, the viewer is sent back to the section menu. The secret link is a short slideshow with no menu, but it has two different music tracks that the viewer can switch between while the video sequence is playing.

To prevent things from getting mixed up, I created the directory structure below to organize the files. The image



A BETTER DATABASE CAN SPEED UP YOUR DEVELOPMENT CYCLE

If your back-end database isn't a good match for your front-end development, you need a new database.

Caché is the *post-relational* database that combines high-performance SQL for faster queries and an advanced object database for rapidly storing and accessing objects. With Caché, no mapping is required between object and relational views of data. That means huge savings in both development and processing time.

Applications built on Caché are massively scalable and lightning-fast. Plus, they require minimal or no database administration.

More than just a database system, Caché incorporates a powerful Web application development environment that dramatically reduces the time to build and modify applications.

The reliability of Caché is proven every day in "life-or-death" applications at thousands of the world's largest hospitals. Caché is so reliable, it's the leading database in healthcare – and it powers enterprise applications in financial services, government and many other sectors.

We are InterSystems, a specialist in data management technology for twenty-five years. We provide 24x7 support to four million users in 88 countries. Caché is available for Windows, OpenVMS, Linux and major UNIX platforms – and it is deployed on systems ranging from two to over 10,000 simultaneous users.

InterSystems
CACHÉ™
Make Applications Faster

Try a better database. For free.

Download a free, fully-functional, non-expiring version of Caché or request it on CD at www.InterSystems.com/match3

© 2003 InterSystems Corporation. All rights reserved. InterSystems Caché is a registered trademark of InterSystems Corporation. 10-03

directory eventually will hold the completed DVD, while the raw photos go in the photos/setN directories and the video files go in the titleN or main directories:

```
dvd
        - title1
        - title2
        - title3
        - title4
        - title5
        - mainmenu
        - photos
                - set1
                - set2
                - set3
                - set4
                - set5
        - image
```

Processing Pictures

In order to make the slideshows, we need to transform the pictures into video clips. The mjpegtools suite includes a utility called jpeg2yuv that transforms a JPEG file into YUV (a color model used for video pictures) format, which then can be piped into mpeg2enc to generate an MPEG-2 video file. Before this transformation happens, though, we need to make sure the picture files are formatted correctly for DVD. It is advisable to alter your pictures manually to fit the PAL or NTSC resolution constraints (720×576 and 720×480, respectively) before feeding them into jpeg2yuv. mjpegtools includes a utility called yuvscaler that automatically resizes incoming YUV data into the correct resolution, but this can alter the aspect ratio of your pictures and cause distortion.

A sample command line to generate a slide looks like this:

```
jpeg2yuv -n 125 -I p -f 25 -j picture1.jpg | \
mpeg2enc -f 8 -o slide.mpg
```

The -n option tells jpeg2yuv how many frames it should generate, in this case 125 (5 seconds at 25fps); you should use 29.97 for NTSC. The -I option tells it to use progressive interlace (rather than throwing parts of the image away), and -f sets the required frame rate. The -f option to mpeg2enc tells it to create a DVD MPEG-2 video stream. This process needs to be repeated for all the photos to be included on the DVD, so you might want to write a quick Perl or shell script to automate this process.

Adding Music

Now we have video clips of separate slides. The next step is to add a music backing track that plays across each clip. The DVD specification allows for the use of PCM, AC-3 or MPEG-2 audio, at various bit rates. For the photo album, I decided to use MPEG-2 audio; I decided that I didn't need the higher audio quality of AC-3 or PCM. mjpegtools includes a tool called mp2enc that encodes WAV files in MPEG-2 format. Another tool from mjpegtools, mplex, multiplexes one or more audio files into a video MPEG stream. To do multiple audio files, you need the CVS version of mjpegtools. Start by concatenating all the video files required for a particular slideshow using **cat**, as mplex seems to have problems with handling multiple input files:

```
cat *mpg > video.mpg
```

In the title5 directory (the fifth title is going to have two audio tracks), I ran mplex on this new file:

```
mplex -f 8 -o video%d.mpg photos/set5/video.mpg \
audio1.mp2 audio2.mp2
```

This creates a series of files called video1.mpg, video2.mpg, ..., videoN.mpg in the current directory, multiplexed with the two audio tracks, audio1.mp2 and audio2.mp2. The -f 8 argument to mplex ensures that the new MPEG files are DVD-compatible, the same as mpeg2enc.

One issue to be aware of is that mplex multiplexes the whole audio track. If the track is longer than the combined length of the video clips, the final clip is extended to cover the rest of the audio. You may want to adjust the frame times in jpeg2enc phase or edit the audio track to prevent this from happening.

Making Menus

The process for creating DVD menus is similar to generating the slides. Using The GIMP, make a new image of size 720×576 (for PAL), with a resolution of 75dpi in the x-axis and 80 in the y-axis (NTSC values are 81 and 72). Add an alpha channel, then create the menu picture you desire (see Figure 1 for the base image of the main menu). Once this is finished, add another layer and mark out the location of the buttons that can be highlighted, as seen in Figure 2. A maximum of four colors can be used for this mask; I used red to mark the button areas, and white to make sure that the text still was visible when a button was selected. When you're satisfied with the menu, export the background layer as a JPEG file and the mask as an indexed PNG. Make sure you've set the number of colors to four, three if you're saving the transparent background as a color. This process is repeated for all menus

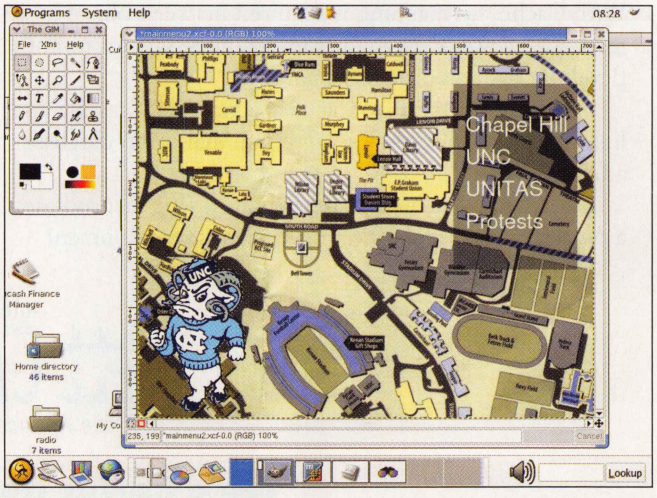
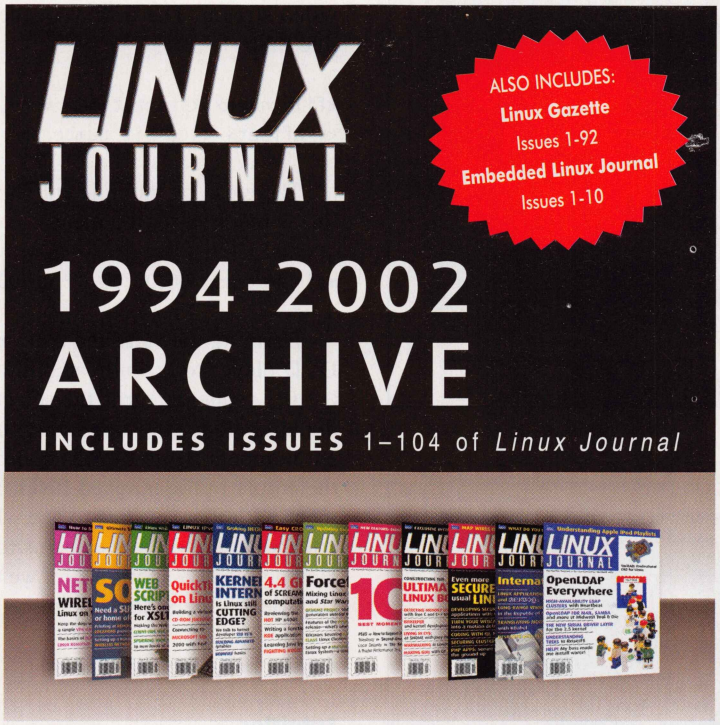


Figure 1. The Main Menu Image

Now Available!



Order your newly updated, easy-to-use 1994-2002 Linux Journal Archive CD-ROM before January 1, 2004, and save 10%!

http://store.linuxjournal.com/

Regular price is $29.95, sale price of $26.96 available until January 1, 2004.

required in the DVD (see Figures 3 and 4 for a further example).

As before, use jpeg2yuv and mpeg2enc to create an MPEG file, and mplex to multiplex it with an audio file. For some reason, menus need an audio file to work, so if you want silent menus, you have to use a short burst of silence. Buttons in DVD are implemented as part of the subtitle stream, so we use dvdauthor's submux tool to define a menu's buttons. The program reads in subtitle information from a text file and multiplexes it into the video stream. The format of the submux files looks like this:

```
<file_name> <start_time> <end_time> <x_offset> \
<y_offset> <4 8-bit numbers>
```

There can be multiple entries in a .sub file, but only one line is required for the menus:

```
mainmask.png 00:00:00.00 00:00:00.00 0 0 0 1 0 1
```

Defining the start and end times as zero instructs the DVD player to keep the subtitle on screen continuously. The 0 1 0 1 sequence turns the selected button red when using the default palette file supplied with dvdauthor. The four different numbers (from 0–255) control the transparency values of the colors in the indexed PNG. You might want to experiment with these values to see what effects you can create, but the sequence above seems to produce reasonable results. Having made a subtitle file, we then run submux to add the subtitles to the video stream:

```
submux menu.sub < video.mpg > menu.mpg
```

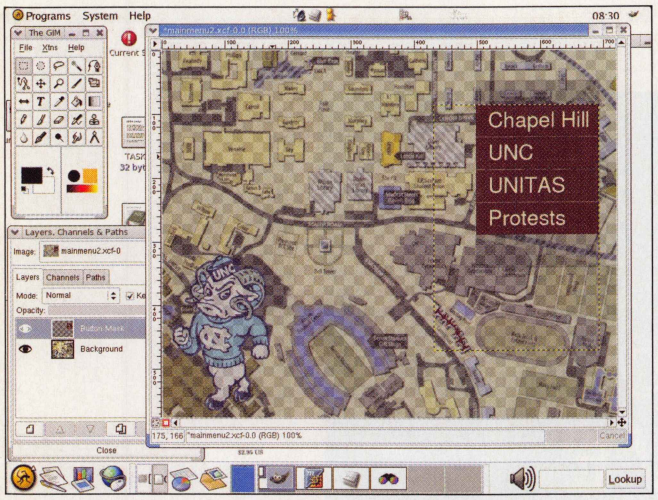


Figure 2. Making the Main Menu Mask

Figure 3. A Submenu with Thumbnail Previews

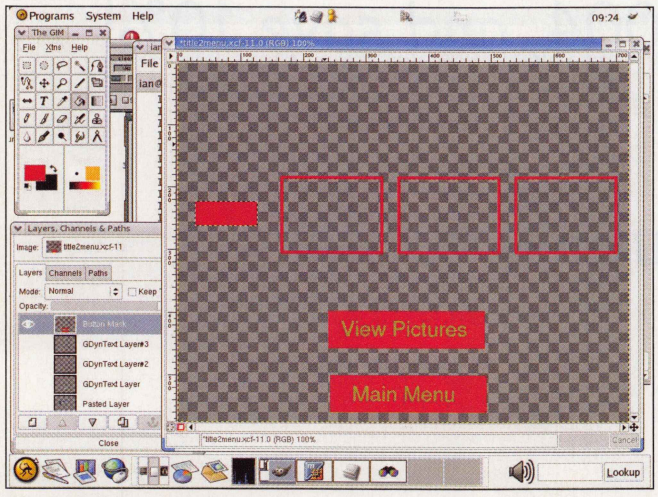


Figure 4. The Mask for the Submenu

Subtitles

As previously mentioned, a DVD can have up to 32 different subtitle streams in a VTS. Although they're called subtitle streams, they actually are implemented as overlaid picture files rather than as a text representation, so they can do other things as well. An example of this is the "White Rabbit" feature seen on *The Matrix* DVD. Subtitling is fairly easy, as it's essentially the same process as creating menus, only without the background layer. Figure 5 shows an example subtitle picture, providing a title for a photo. A sample submux description for this

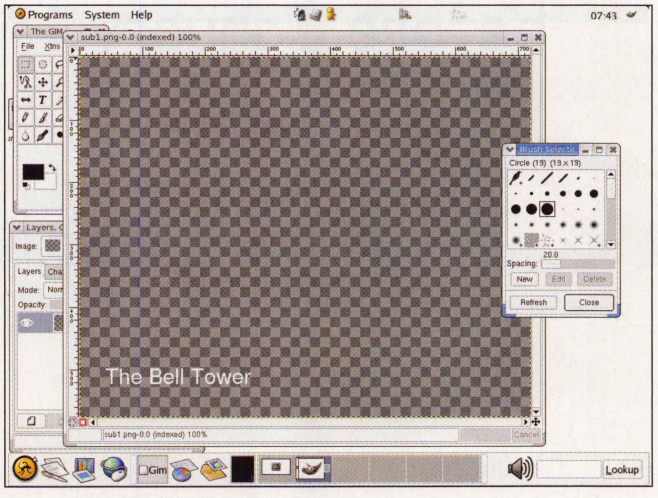


Figure 5. A Subtitle Image

Polywell 64-bit Computers
Superior performance for mission critical business applications

POLY SYSTEMS

800.765.9686

www.Polywell.com/us/LJ

15 YEARS OF CUSTOMER SATISFACTION

5-Year Warranty, Industry's Longest
First Class Customer Service

Only AMD64 technology runs today's 32-bit and tomorrow's 64-bit software

The ONLY Windows-compatible 64-bit PC processor

DVD Authoring and WEB Streaming

POLY 900VF-3200

AMD Athlon™ 64 3200+ Processor
with Hyper Transport™ Technology
1600MHz System Bus
1GB **400MHz DDR** memory (Exp. to 4GB)
120G Ultra ATA 7200RPM Drive
52x52x24x CDRW, 16X DVD-ROM Drive
NVIDIA GeForceFX™ 128M 8X AGP Graphics
19" High-Res.25mm Graphics Monitor
Dolby **5.1 Sound+Subwoofer** Speakers
Wireless Keyboard, Wireless Mouse
100Mbit LAN, Fax/Modem, 1394 Firewire
MS® Windows® XP Home, Lotus SmartSuite
10-Bay Silver Mini-Tower w/350W P/S

$1,599

Order# **900VF-3200SP0909A**

POLY 900NF3-FX

AMD Athlon™ 64 FX-51 Processor
with Hyper Transport™ Technology
1600MHz System Bus
1GB **800MHz DDR** memory (Exp. to 8GB)
200G Ultra ATA 7200RPM Hard Drive
52x52x24x CDRW, 16X DVD-ROM Drive
NVIDIA GeForceFX™ 5900 128M 8X AGP
19" High-Res.25mm Graphics Monitor
Dolby **5.1 Sound+5.1 Subwoofer** Speakers
Wireless Keyboard, Wireless Mouse
100Mbit LAN, 1394 Firewire
MS® Windows® XP Home, Norton Anti-Virus
10-Bay Aluminum Tower w/400W Quiet P/S

$2,450

Order# **900NF3-FX51-0909B**

POLY 900NF3-DV

AMD Athlon™ 64 FX-51 Processor
with Hyper Transport™ Technology
1600MHz System Bus
1GB **800MHz DDR** memory (Exp. to 8GB)
400G (2x200) Ultra-ATA 7200RPM 8M Cache
Sony DVD-RW, 52x52x24x CDRW Drive
NVIDIA QuadroFX™500 8X AGP 128M OpenGL
19" High Res. .21mm Trinitron Monitor
TV Turner + TiVo w/DVD,Video Editing
1394 Video In/Out, **Wireless** KB+Mouse
Dolby **5.1 Sound+5.1 Subwoofer** Speakers
Windows® XP Professional, Norton AntiVirus
10-Bay Aluminum Tower w/400W Quiet P/S

$2,999

Order# **900NF3-FXDV-0909C**

Polywell Computers, Inc 1461 San Mateo Avenue South San Francisco CA 94080 650.583.7222 Fax: 650.583.1974

AMD and ATHLON are trademarks of Advanced Micro Devices, Inc. All other brands, names are trademarks of their respective companies.

title that keeps it onscreen for two seconds would be:

```
sub1.png 00:00:00.00 00:00:02.00 0 0 0 255 0 255
```

In this application, I'm working with separate, short MPEG-2 clips, so each has to have a separate submux description. For longer clips, a submux file probably would have multiple entries.

Putting It All Together

Having made the video slides, subtitling them and creating the menus, the only thing left to do is assemble everything together using dvdauthor. dvdauthor works in two modes, one for normal titles and another to define the VMGM menu; the VMGM can't have any video information apart from the menu. Let's start with the first title:

```
dvdauthor -o tmp -m -P \
-b 239x397-489x457,subtitle32+vts1 \
-b 239x500-489x560,vmgm1 \
-b 27x223-127x263,subtitle32+vtsm.2 \
-b 165x184-325x305,subtitle32+vts1.1 \
-b 352x184-512x305,subtitle32+vts1.2 \
-b 539x184-699x305,subtitle32+vts1.3 \
\
title2/titlemenu1.mpg -m -P \
-b 239x397-489x457,subtitle32+vts1 \
-b 239x500-489x560,subtitle32+vmgm1 \
-b 27x223-127x263,subtitle32+vtsm.1 \
-b 165x184-325x305,subtitle32+vts1.4 \
-b 352x184-512x305,subtitle32+vts1.5 \
-b 539x184-699x305,subtitle32+vts1.6
\
title2/titlemenu2.mpg -t -P \
\
title2/v1.mpg title2/v2.mpg title2/v3.mpg \
title2/v4.mpg title2/v5.mpg title2/v6.mpg \
 -i post=3Dvtsm
```

This code defines two title menus with six buttons each, six MPEG video clips (dvdauthor defines the transition from one clip to another as a chapter point) and a post-video instruction that returns the DVD player to the title menu. The coordinate system is the same as The GIMP's, so you can use that to get the necessary regions. A button can perform various actions (see `dvdauthor --help` for a list of possibilities). In the example above, the first button sets the audio track to zero, sets up the required subtitle stream (stream 0 is bizarrely numbered 32, stream 1 is 33 and so on) and plays the video associated with the title. The second button returns the player to the main VMGM menu, and the third button moves the DVD player to the second title menu. The other three buttons move to specific chapters within the video stream, corresponding to clicking on one of the thumbnail pictures. The -o option is for an output directory, in this case the image/ subdirectory. The program works incrementally, so running the same command line twice creates an additional title set rather than updating the original.

For the main menu, you need to pass the -T option to dvdauthor so it knows it should create the required VMGM information. Then, link in the VTS files it already has created:

```
dvdauthor -o tmp -T -m \
-b 497x89-693x136,vtsm1 \
-b 497x138-693x187,vtsm2 \
-b 497x189-693x239,vtsm3 \
-b 497x240-693x289,vtsm4 \
-b 426x405-490x474,vts5 \
-P mainmenu/mainmenu.mpg
```

Testing and Burning

It's a good idea to test the DVD image before burning it onto disc. This can be done using xine; all you need is to give xine an argument like this:

```
xine dvd:/path_to_dvd_root/VIDEO_TS/
```

and xine should act as though it is playing the information

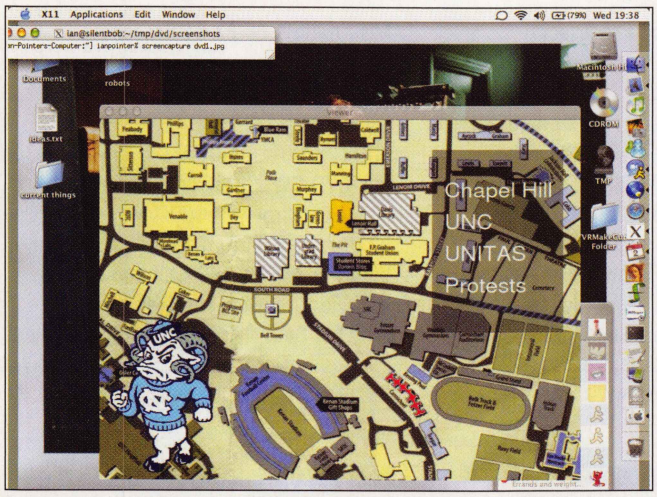


Figure 6. The DVD Running on My iBook, with the Secret Link Selected

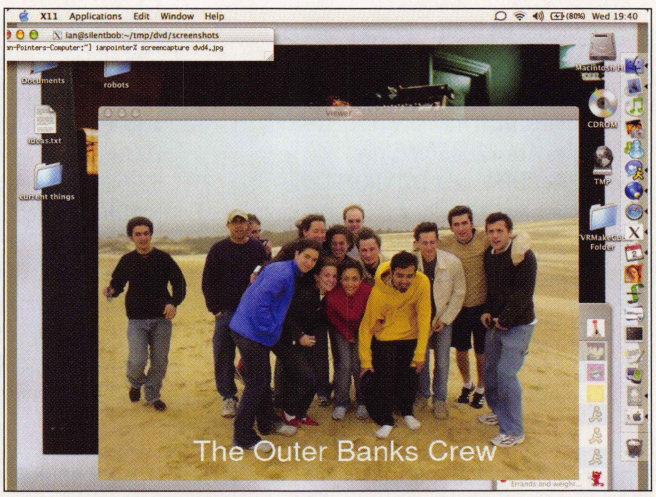
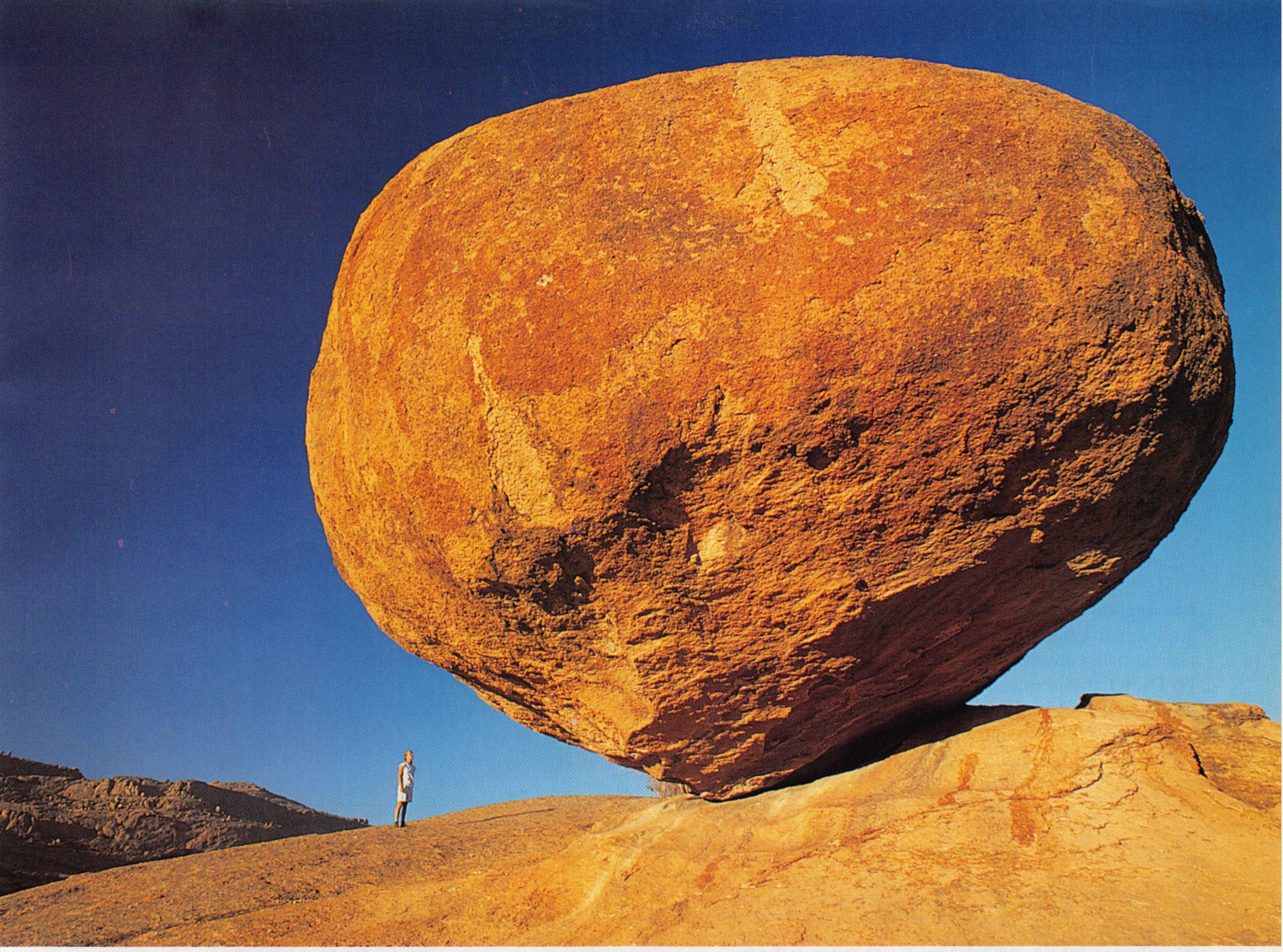


Figure 7. One of the Slides Running on My iBook's DVD Player



Network backup seems insurmountable if you don't have the right solution.

Introducing Arkeia 5.2
The right solution.

If you're responsible for protecting your company's heterogeneous network, the Arkeia 5.2 Linux or Unix backup server solution is just what you've been looking for.

Arkeia has always been noted for powerful, scalable, reliable and cost-effective backup solutions. Now we've added even more options to give you greater control and ease of use.

New features, greater benefits
Our completely redesigned reporting functions allow you to produce highly customizable machine-based reports as well as more developed messages that inform you about successful operation completion.

Quantum DLTSage™ support for SDLT tape devices helps you predict and prevent storage system errors.

An optimized catalog algorithm lets you dramatically increase backup speed for large file servers—five times faster for 10,000 files within a single directory and up to 50 times faster for 150,000 files!

We've also expanded GUI capabilities for maximized security. You can easily clone tapes for off-site storage. Combine this with Arkeia's Disaster Recovery module, and you can automatically rebuild servers from scratch if they are destroyed or stolen.

In addition, a new hot backup plug-in is available for MS-Exchange that complements our online backup solutions for Oracle, Lotus, DB2 and MySQL.

Try Arkeia 5.2 for 30 days. Free!
The best way to prove that Arkeia 5.2 is the right solution for you is to let you try it—**FREE**—for 30 days. We'll even include free installation tech support. Simply download the demo version at www.arkeia.com.

arkeia®
ENTERPRISE NETWORK BACKUP

www.arkeia.com

from a DVD. As dvdauthor is incremental, you should be able to use xine after creating each separate title to ensure that you're doing things properly.

Once you're happy with your DVD, it's time to burn. I used the cdrecord-prodvd application for DVD-R burning. The operation is the same as cdrecord, so first you need to create an ISO image using mkisofs:

```
mkisofs -o <output_filename> -dvd-video \
<path_to_dvd_root>
```

Then use cdrecord.prodvd to burn it:

```
cdrecord.prodvd dev=3D0,0,0 -pad -dao \
<path_to_DVD_image>
```

Figures 6 and 7 show the DVD running on my iBook.

Conclusion

Although dvdauthor doesn't have the easy-to-use interface of professional applications, it provides all you need to make DVDs to the same standards as Hollywood uses, merely for the price of DVD-R media. Hopefully, this article has shown you the basics of DVD authoring and provided you with some ideas for your own applications. Get creating!

Ian Pointer is an unemployed computer science graduate in the UK who has far too many DVDs in his house and now plans to make more. He can be reached at ian@snappishproductions.com.

RESOURCES

cdrecord.prodvd: www.fokus.gmd.de/research/cc/glone/employees/joerg.schilling/private/cdrecord.html

dvdauthor: dvdauthor.sf.net

dvdrtools: www.nongnu.org/dvdrtools

The GIMP: www.gimp.org

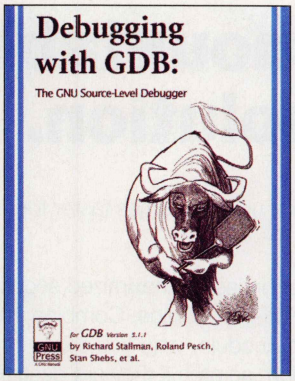
mjpegtools: mjpegtools.sf.net

xine: http:/www.xinehq.de

The Definitive GCC Reference Manuals

GNU Press

GNU PRESS ▪ 617-542-5942 ▪ www.gnupress.org ▪ Free Software Foundation

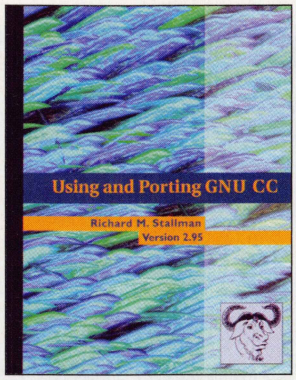


Debugging with GDB

for GDB version 5.1.1

ISBN 1-882114-88-4
346 pages • $25

Debuggers making you batty? This top selling tutorial is designed so the reader can utilize GDB after finishing the first chapter. Finish the book and you will be a master!

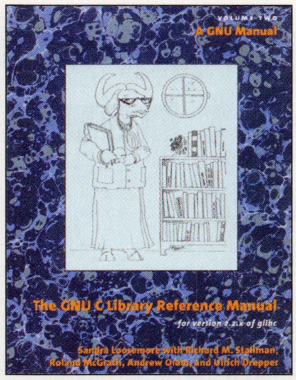


Using and Porting GNU CC

Version 2.95

ISBN 1-882114-38-8
588 pages • $35

A thorough reference guide to installing, running, debugging, configuring and porting GCC. Covers C, C++, Objective C and Fortran front-ends.

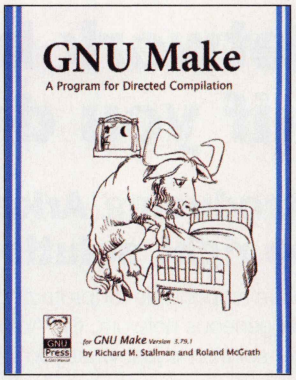


The GNU C Library Reference Manual

for version 2.2.x of glibc

ISBN 1-882114-55-8
2 vol. • 1,275 pages • $60

A comprehensive guide to implementing the standard C libraries. Contains code examples and usage recomendations. Useful for system admininstrators and programmers.



GNU Make

for GNU Make Version 3.79.1

ISBN 1-882114-82-5
196 pages • $20

This best selling tutorial is indispensible to free software system maintainers. Written by the program's authors, it also includes an introductory chapter for novice users.

Cool. Fast. Reliable.

Launch a Microway® Linux cluster and solve your problems in life sciences, structure design, weather, astronomy . . .

Run cool. Microway's **CoolFlow**™ 1U chassis is designed to provide the best cooling in the industry. Low temperatures result in higher CPU and memory reliability, plus longer cluster life.

Take control. Monitor *and* control your cluster remotely via our secure web-based GUI. Microway's **NodeWatch**™ hardware and **MCMS**™ cluster management software report on temperatures, voltages, and chassis fans, and let you set thresholds for notification and/or failsafe shutdown if problems occur.

Today's HPC market is wilder than ever. Microway's expert design team will guide you through the whitewater of hardware, software, and storage solutions. Our clusters are competitively priced and delivered on time—and their proven reliability yields low total lifetime cost. What's more, our tech support is legendary.

Call us to test drive **NodeWatch**™ online, arrange a benchmark, and discuss your next Linux cluster with people who speak your language. Visit our website for *HPC Times* technical news.

CoolRak™ Cabinet Features Xeon XBlade™ Technology

- 1.056 Teraflops per 44U cabinet running 3.06 GHz Xeons
- NodeWatch™ remote control and monitoring
- CoolFlow™ proprietary 1U and 4U chassis
- Myrinet, GigE, or InfiniBand™ connectivity

GSA# GS-35F-0431N

intel inside ITANIUM 2 AMD Opteron intel inside XEON

Microway

Scientists have been counting on us for over 20 years.℠

508.746.7341 • www.microway.com

Managing Audio with Pd

Ready to turn your Linux box into the ultimate audio effects machine? With the Pd environment, you can build reverb and more, visually. BY PETER TODD

Pure Data (Pd) is a real-time visual programming environment for audio and other multimedia applications. With it you can make patches that perform operations on audio and visual data. These patches are represented visually; you "draw" where you want the signal data to go and what you want to happen to it. This process is similar to how you program modular analog synthesizers. The process also is well suited to how you end up programming sound and video applications; signals come in and go out, and you manipulate various signals along the way. Due to space constraints, this article covers only the more mature audio capabilities of Pd. If you later want to try the video processing plugins, you should find that many of the concepts are similar to those for audio.

Messages and Audio Signals

Pd uses two main types of data, messages and audio signals. Messages are sporadic, like MIDI note events. They can contain numbers or strings and are used to pass around information, such as "set the gain on the output to *x*". Audio signals are constant; whenever the DSP code is turned on, audio is being transferred. Internally in Pd, audio is represented by 32-bit floating-point numbers. This means that unlike conventional analog or digital sound processing, Pd signals can have nearly any amplitude you want. During processing you can make a audio signal really quiet for one stage and amplify it for another, with no loss in quality. Of course, when the signal is sent to a hardware output, you must make sure it's within the usable range of –1 to 1, or the audio gets clipped.

These messages and audio signals are manipulated by various types of boxes, described below. When put together, this collection of connected boxes is called a patch.

Boxes—Getting Some Work Done

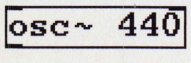


Figure 1. An Object Box

Boxes do all the work. Pd has four main types of boxes: object, message, GUI and comment. These boxes perform operations on messages and audio, provide ways to give user input and document what's been done. Object boxes in turn are divided into two types, control objects and tilde objects. Control objects work with messages and therefore perform their functions sporadically. Tilde objects work with audio data and perform their functions constantly.

Message boxes send their contents to their output port when the user clicks on them or when they receive a message on their input.

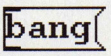


Figure 2. A Message Box

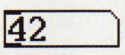


Figure 3. A GUI Box

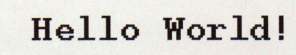


Figure 4. A Comment Box

GUI simply refers to boxes you can interact with, such as the number box on the left.

Finally, comments allow you to put text into a patch. They actually don't affect anything.

Starting Pd and Making Your First Patch

Assuming you've compiled and installed Pd by now, you should try to start it. First, make sure you've set the setuid bit on the Pd executable, and make sure it's owned by root. Although this could be a security risk, you need to do this to enable real-time scheduling if you want to run Pd as any user other than root. If you don't and real-time scheduling isn't activated, you'll hear a lot of clicks and pops whenever any other process, even the X server, tries to do anything.

Run Pd with the -rt option and any other options you need in your setup. I'd recommend using -verbose, because Pd itself is not overly chatty and the verbose option does provide some useful information. When that's done you should see a window similar to the one shown in Figure 5. The IN and OUT boxes

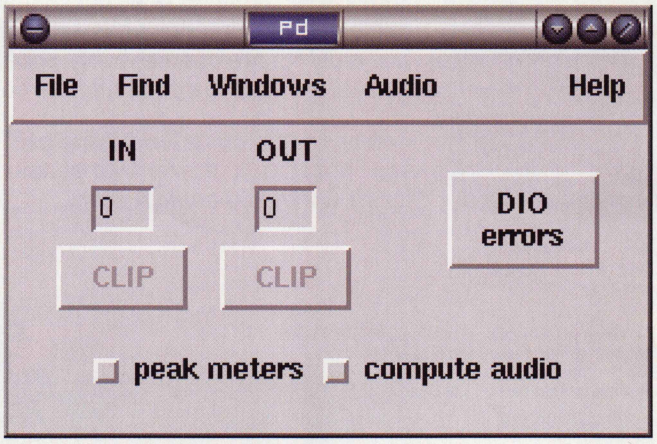


Figure 5. The Main Window

are peak meters for input and output, respectively, and can be enabled by clicking the peak meters option. If either clips, the respective CLIP box will go red. The DIO errors button flashes if there are any synchronization errors in the input or output. Click on it to see a list of recent errors. Finally, the compute audio check box turns audio processing on or off.

First, let's create a new blank patch in which to work (File→New). From this we are going to create a simple patch to print "Hello World!" to standard output. So, we need a message box to hold the message "Hello World!" and an object box to do the printing. Both of these can be created from the Put menu. You also can use the accelerator keys: Crtl-1 to place the object box and Crtl-2 to place the message box. Once you've done that, to enter the right text in the boxes, click on them and type. You also need to connect the outlet port on the bottom of the message box to the inlet port on the top of the object box. Your patch should look like the one shown in Figure 6. Get out of edit mode by pressing Crtl-E. Now try clicking on the "Hello World!" message box; if all goes well, a message saying so is printed to the terminal in which you started Pd.

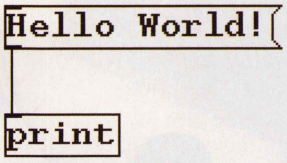


Figure 6. The Standard "Hello World!" Example

Using Messages

Now, let's try a more complex example, making a patch that adds two numbers together and displays the result. Make a patch like the one shown in Figure 7, using two number boxes at the top, an object box in the middle and a number box on the bottom. Next, exit edit mode (Crtl-E again), and try changing the numbers at the top. You can do this by clicking and typing or clicking and dragging; move up to increase the number and move down to decrease it. As you've probably noticed, changing the number on the left makes the sum immediately change, but changing the number on the right does nothing. Why?

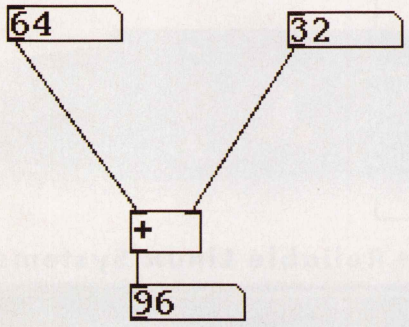
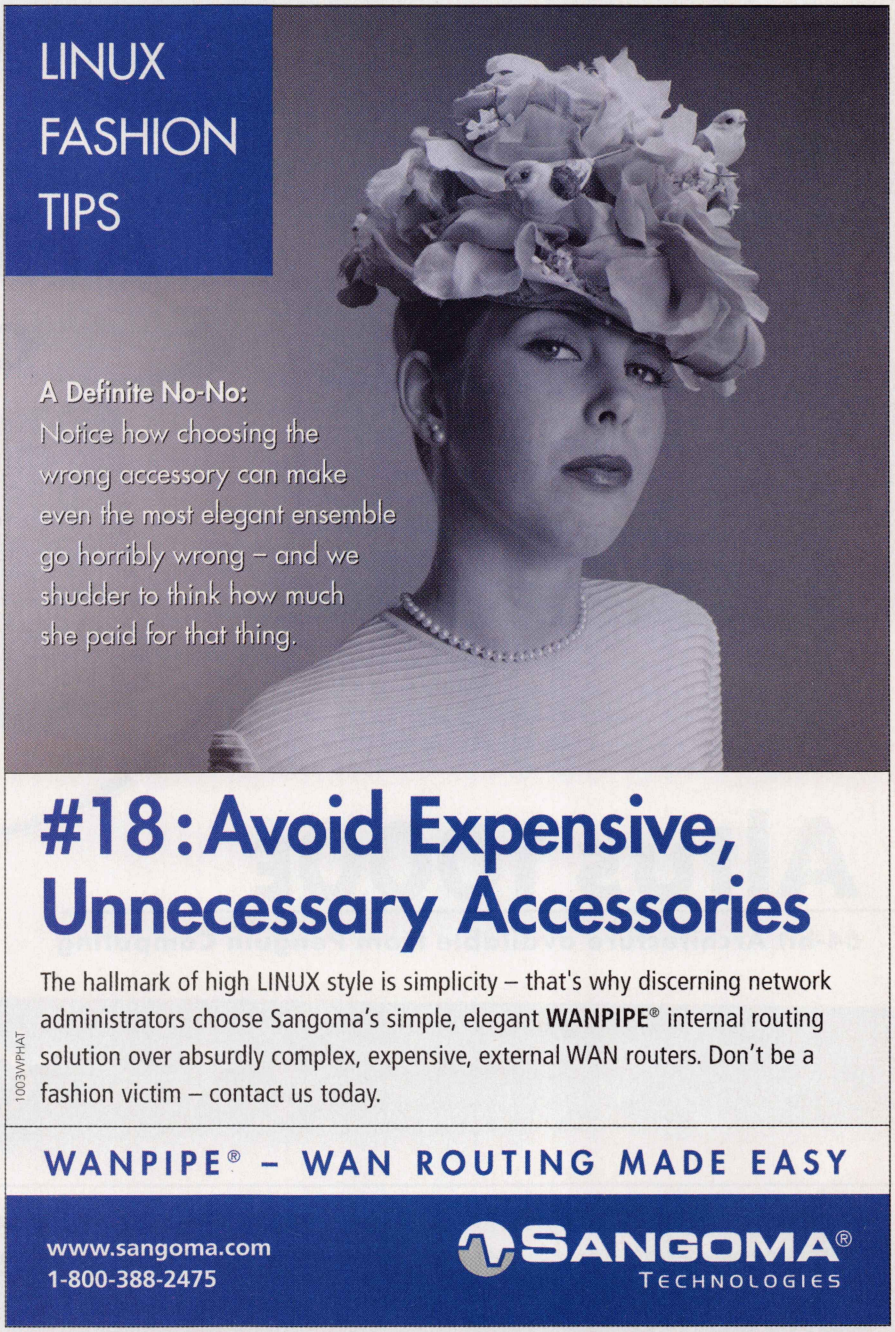


Figure 7. Adding Two Numbers

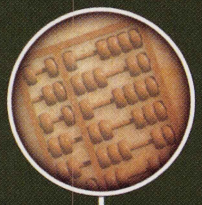
Nearly all objects in Pd treat their left-most inlet as hot, meaning that any changes in their value affects an immediate change in the output. The other inlets are cold. Changes in their values don't trigger any change in the output. The new value simply is put into storage until a computation is triggered by the hot inlet, at which point the new value is used.

But, what if you do want a change in a cold outlet to trigger a change in output? One way to accomplish this is to inset a message box. In Figure 8, I connected the outlet of the number box on the right to the inlet of a message box. The outlet of that message box then is connected to the hot inlet of the addition



EVOLUTION of COMPUTING

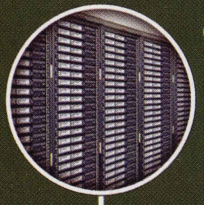
abacus is used



personal computer is here

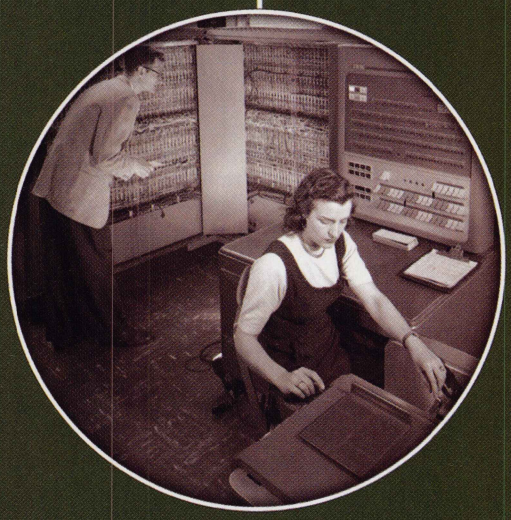
www takes hold

present time



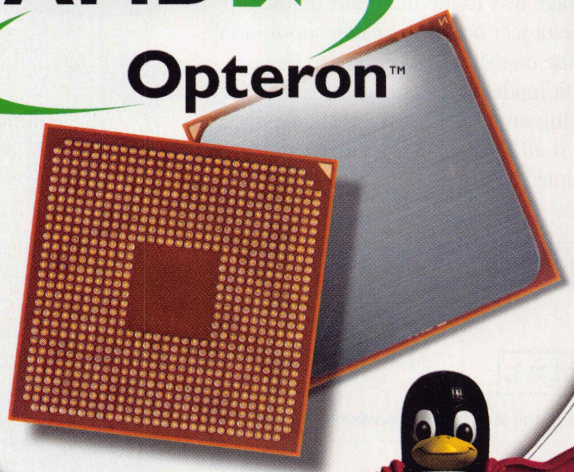
64-bit Architecture
Clustering Power

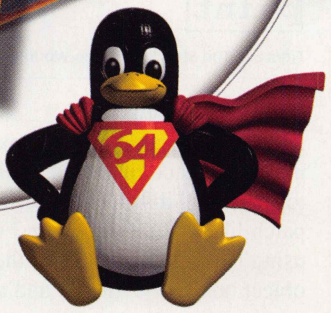
The Age of Thought | The Age of Technology | The Age of Penguin Computing



commercial computing arrives

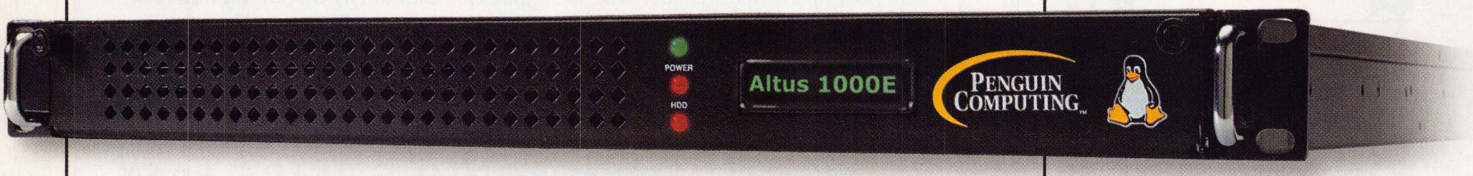
AMD Opteron™





Altus 1000E

64-bit Architecture available from Penguin Computing



The World's Most Reliable Linux Systems

PENGUIN COMPUTING™

1-888-PENGUIN

www.penguincomputing.com

©2003 Penguin Computing. All rights reserved. AMD, the AMD Arrow logo, AMD Opteron, and combinations thereof, are trademarks of Advanced Micro Devices, In

object box below it. When message boxes receive any message at all on their inlets, they send all their contents as a new message on their outlets. So when the right-most number box is changed, it sends a message to the bang message box, which then sends a bang message to the addition object box. Bang messages mean "Do something!", so any object box receiving one on its hot inlet immediately performs whatever computation it's told to do. We use this behavior here to make our addition object box act as though it has two hot inlets.

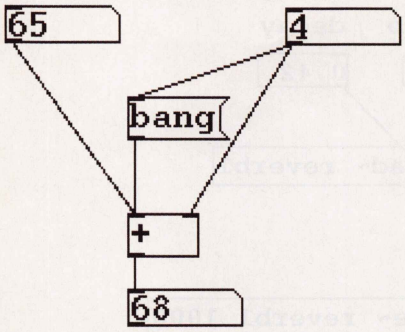


Figure 8. The Naïve Way to Deal with Hot and Cold

But wait, doesn't that mean the bang message has to arrive after the number? If it doesn't, the patch won't work, right? Well, yes; depending on the order in which you made your connections, you already might have noticed that it doesn't. Indeed, in Figure 8 the numbers don't add up precisely because of this problem. So, we need a way to ensure that the bang message arrives after the number. An easy way to do this is to insert a delay, as shown in Figure 9. Interestingly, a delay of 0 actually works. The message simply is delayed by one DSP cycle, thus ensuring that the bang message arrives second.

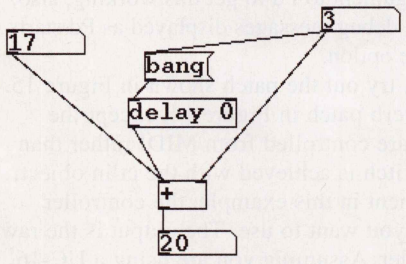


Figure 9. Dealing with Hot and Cold the Right Way

Audio Processing

The most basic audio function is input and output. The adc~ tilde object, standing for analog-to-digital converter, performs the first task; and the dac~ object, digital-to-analog converter, performs the second. Both objects operate on the first two channels by default. If you want to change this—for instance, if you have a multichannel sound card such as a Hammerfall HDSP—you can enter channel numbers as arguments, and the respective channels then are mapped to their respective inlets or outlets. Figure 10 shows a simple example where stereo input is flipped and routed to output. Because stereo input is the default, the channel numbers in the example are redundant, but we've included them anyway for demonstration purposes.

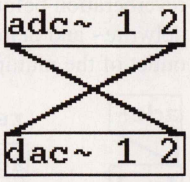


Figure 10. Flipping Left and Right in a Stereo Signal

Sound data is a sequence of numbers at a specific sample rate, so it's possible to apply arithmetic operators to sound data. All you have to do instead is add a tilde to the end of the operator you want to use. For instance, Figure 11 turns a stereo signal into a mono one by adding the left and right channels together. Another useful operator is multiply, *~, which acts as a gain control. Remember, though, signals clip when output to hardware if they are beyond the values –1 and 1.

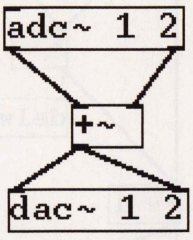


Figure 11. Turning Stereo into Mono

Figure 12 shows a more complex example. You might want to turn your speakers down some before you run this one, because it's loud. First, the osc~ object at the top is a sine-wave generator, in this case running at 440Hz. This signal is split into two, the left side going directly to the addition and the right side first going through a multiplication.

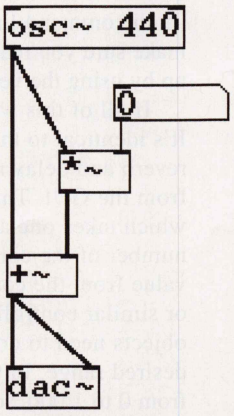
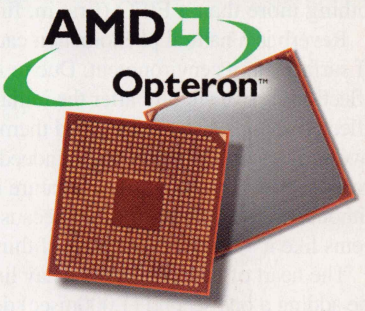


Figure 12. Using Inversion to Cancel Two Signals

Now try

the
ALTUS 1000E
incorporating AMD
Opteron™ Processors

- 32-bit and 64-bit Operation
- Dual AMD Opteron™ 200 Series Processors
- Up to 16GB DDR RAM
- Dual Gigabit Ethernet Ports
- Up to 400GB IDE Storage Capacity



Leverage exciting new 64-bit cluster critical features, such as large memory addressability for data-intensive applications yielding optimal cluster performance.

Enable High Performance Cluster customers to easily transition to 64-bit computing applications at their own pace, without sacrificing performance on the immense installed base of current 32-bit applications.

Provide lower Total Cost of Ownership (TCO) and long-term IT investment protection as customers transition to 64-bit computing on the same hardware they already own.

1-888-PENGUIN

PENGUIN COMPUTING™

www.penguincomputing.com

entering –1 into the number box. The sound stops, why? If you remember your wave physics from high school, you know that waves can cancel one another out. In this case, the –1 creates a perfect inverse of the original signal. So where the original is at 1, the inverse is at –1. When you add these two signals together you get 0, silence. You also can try holding down the Shift key and dragging on the number box; the numbers should change slowly enough that you can hear the sound getting quieter as you approach –1 and then finally stopping.

Making a Simple Reverb Filter

Let's put these components together and try to do something useful with them; let's make a simple reverb filter, using the above techniques and one new one, the delay line. If you already have experience with designing sound effects, you probably know what a delay line is; if not, think of it as a buffer for sound. Sound that goes into a delay line inlet is stored for a fixed period of time before it comes out again. All sorts of complicated technologies have been used to implement delay lines in the analog realm, some involving special springs and loops of tape. Fortunately for us, we are dealing with digital signals where a delay line can be implemented with nothing more than a FIFO (first in, first out) buffer.

Reverb is a natural phenomenon caused when a sound reflects off surfaces in its environment. Due to increased distance, these reflections arrive slightly after the initial sound is heard. These reflected sounds also are reflected themselves. Most environments have some degree of reverb, and indeed a carefully controlled level of reverb is a sought-after feature in concert halls. A room without reverb is discomforting, because every sound you make seems like it's been snatched out of thin air.

The heart of this filter is the delay line. Figure 13 shows a delay line adding a one second (1,000msec) delay to the sound input. Delay lines in Pd are named, meaning that to use one you need two separate objects, a writer and a reader. The writer, delwrite~, takes two arguments: the first is the delay line's name, and the second is the maximum delay you want to use in msec. You can have only one writer per delay line. The second half is the reader, delread~, which also takes two arguments: the name of the delay line and amount of delay you want. Unlike the writer half, you can have as many readers as you want, with any delay time you want. An additional type of reader, the variable delay object (vd~), can change its delay in response to an audio signal, but that's beyond the scope of this article.

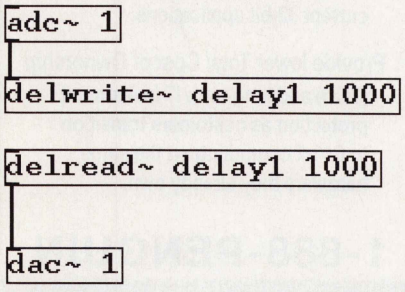


Figure 13. A Delay Line

To recap, a reverb filter needs one straight signal path, the main signal, and one or more delayed signal paths, the reverberations. The reverberations are recursive, so some form of feedback needs to be in the system. Finally, although our inputs and outputs are stereo, in the interest of space we simply implement a mono reverb filter. For this, the incoming left and right need to be mixed together and split at the other end. Figure 14 shows such a filter. If we follow the signal patch from the inputs, the first thing you should notice is the *~ 1 object. This object is included to make the patch a bit cleaner. It multiplies the signal by one, which is useless in itself, but it also gives us an inlet to connect both the left and right channels and mix them into one. You can see a similar structure at the bottom of the patch. The reverb signal then is sent to another multiplier to attenuate the reverberation before it goes through another reverb cycle. The delwrite~ and delread~ form a loop of reverberation. Finally, the outlet of the multiplier is sent back to the dac~.

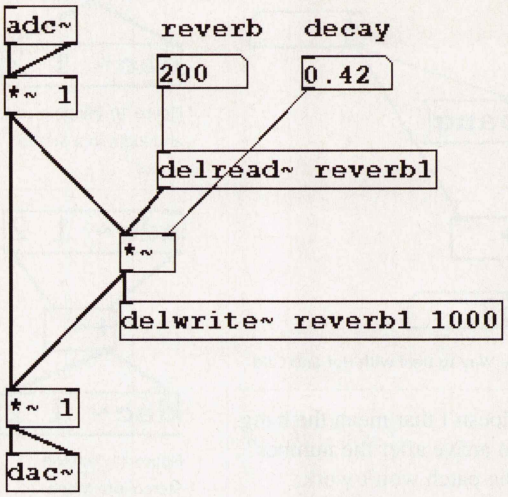


Figure 14. A Simple Reverb Filter

Adding MIDI Control

Make sure the MIDI device you want to use, preferably a control device such as the Evolution UC-16, currently is working. An easy way to verify that MIDI events are being received by Pd is to go to the Test Audio and MIDI menu item in the Help menu on the upper-right corner of any window. Open them and generate some MIDI events. If your controller is working and sending out control messages, as opposed to note events, you should see the numbers under the ctlin (control in) object at the bottom of the window change. You may need to play with the -midi command-line argument to Pd to get this working; also, make sure you read the debug messages displayed as Pd starts up by using the verbose option.

If all of this works, try out the patch shown in Figure 15. It's identical to the reverb patch in Figure 14, except the reverb and delay now are controlled from MIDI rather than from the GUI. This switch is achieved with the ctlin object, which takes one argument in this example, the controller number of the control you want to use. The output is the raw value from the controller. Assuming you are using a UC-16 or similar controller, this is a value from 0 to 127. The other objects need to do a little math on this value to get it into the desired range. In the case of reverb, the range we want is from 0 to 1,000, so we divide by 127 to get a value from 0 to 1 and then multiply by 1,000 to get our final value. For decay, a value of 0 to 1 is good, so we need only to divide.

In business since 1989, ASA builds custom computing and storage solutions compatible with open-source OS (BSD, Linux) and also Solaris and Windows. ASA has always enjoyed strong financial stability and a loyal customer base.

ASA --- Custom Servers and Storage

www.asacomputers.com • 866-382-5263

2354 Calle Del Mundo, Santa Clara, CA 95054

For details/inquiries/customization email **linuxjournal@asacomputers.com**

All systems are pre-loaded with any Linux/BSD version/distribution of your choice. On-site warranty, cross-ship options available.

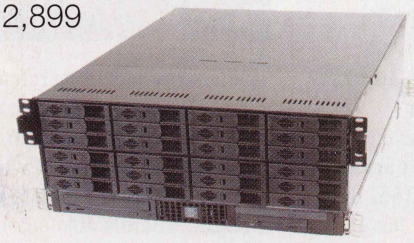
MINI SUPER For Clusters

1U 14" Depth
1 of 2 Intel® Xeon™ processors 2.4 GHz
Serial, VGA, USB 2.0, Mouse, Keyboard
All ports Front Accessible
1 x 10/100, 1 x Gigabit LAN
512 MB DDR ECC (Max 8 GB)
Options: CD, Floppy

$1249

NO-FRILLS STORAGE SERVER

$12,899

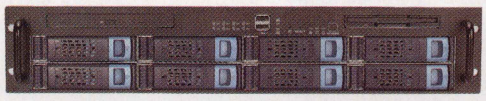


6 TB IDE/SATA storage in 5U !
Dual Intel® Xeon™ processors 2.4 GHz
512 MB DDR ECC (Max 8 GB)
6 TB IDE Raid Storage
Dual Gigabit LAN, CD
Options: Serial ATA Drives, Firewire, DVD±RW, CDRW, 64-bit OS configuration, Additional LAN, Floppy, Fiber Gigabit

8 HOT-SWAP BAYS IN 2U

1 of 2 Intel® Xeon™ processors 2.4 GHz
512 MB DDR ECC (Max 8 GB)
3 x 40 GB IDE in Raid 5
1 x 10/100, 1 x Gigabit LAN
CD, Floppy
Options: SATA, SCSI, Redundant Power Supply, Raid 0,1,5

$1949



Intel®, Intel® Xeon™, Intel Inside® and the Intel Inside® logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
Prices and availability subject to change without notice. Not responsible for typographical errors.

Providing visual feedback on the values to which your effects are set is good practice, so this patch sends a copy of both reverb and decay to number boxes marked as such.

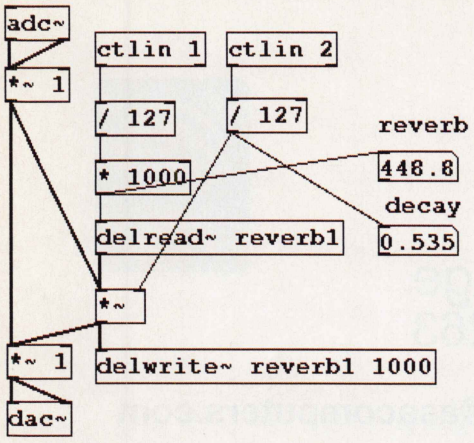


Figure 15. Using MIDI to Control a Patch

There you have it—a fully computerized, MIDI-controlled sound effect made from only a few Pd objects connected together. This is only the beginning of what's possible. Got a MIDI keyboard? You can use it as a control device and make a patch that actually synthesizes the notes. Or, make a patch control the MIDI keyboard. You even could make an entire reprogrammable effects box. Furthermore, you don't have to be content with the existing objects in Pd; you can write your own in C or even in Pd itself. With Linux and Pd, the only limits are your skills and the speed of your CPU.

All of the examples from this article are available from the *Linux Journal* FTP site at ftp.ssc.com/pub/lj/listings/issue116/7062.tgz.

Peter Todd has been using Linux since he was 14. He has a part-time job as the head techie of a small Linux-based sound studio. When he's not working he attends Wexford Collegiate, where he currently is studying ceramics and graphic design.

BOOK REVIEW

Inside the Security Mind: Making the Tough Decisions

by Kevin Day

Prentice-Hall PTR, 2003
ISBN: 0-13-111829-3
$44.99 US

A representative from a leading Irish security consultancy recently gave the following, idiotic advice on one of Ireland's most-listened-to radio phone-in shows: "Install a personal firewall, then sit back and relax—you'll never have to do anything again." If I had been anywhere near this "expert", I would have thrown the book I was currently reading at him. My only regret is it does not come in hardback.

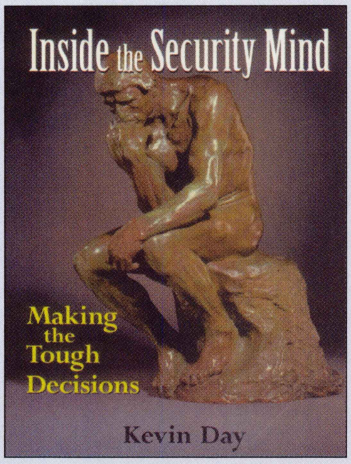
Inside the Security Mind: Making the Tough Decisions by Kevin Day is a must-read security text. Unlike IT security how-to books designed to teach the mechanics, Day's book looks at IT security from a higher perspective, with the emphasis firmly on enabling the reader to think with a security mind. Day's goal is to raise thinking about and awareness of security to a new level.

Day presents the art of IT security in four virtues, eight rules and eight concepts. Rather than drowning in the details of IT security, Day suggests transcending them. For instance and by way of example, it does not matter that you spent 50 hours configuring your firewall and locking it down tight if a user on your network has a modem set up to accept incoming telephone connections.

The first six chapters contain the bulk of Day's original material. The remaining six chapters are more standard IT security fare, including a discussion of various types of attackers, vulnerabilities, targets and exploits. Chapter 8, "Practical Security Assessments", presents the Relational Security Assessment Model, a risk/threat assessment model developed at the author's company. This material is written in a style different from the rest of the book, and I would have preferred that this material, which is the driest in the book, be given the same treatment as the rest. The closing chapters of the book present some discussion of how the earlier ideas can be applied in practice.

If you are looking for advice on securing your brand X router, switch or firewall, you will be disappointed. Day's book is about the bigger picture, and in many respects, he succeeds in presenting exactly that.

Unfortunately, excellent presentation of the material is marred by Day's use



of the term hacker to refer to the bad guys. On page 124 he writes, "I will make life easy and continue the misuse of this term." I would have preferred that he set the record straight. There's also a collection of embarrassing typos that should have been caught by *somebody* before the book went to press. A more extensive index also would be welcome.

These gripes aside, you would be ill-advised to think of yourself as a security expert until you have absorbed this book's message. The first six chapters easily form the basis of an interesting IT security curriculum, so all you academics out there, take note of this title.

—PAUL BARRY

Let the games begin!

You may be running out of pages in your *Linux Journal* calendar, but the contests won't end there! We are pleased to announce that during 2004 we are going to continue to introduce a new contest each month.

Each contest will be listed on the official contest web site: **http://www.linuxjournal.com/contest**, and will also be announced in the *Linux Journal* weekly e-newsletter. (Visit http://www.linuxjournal.com/ to subscribe and make sure you get a head start on each contest!)

You can look forward to more trivia questions, challenges, and exciting prizes, much like these previous contests...

Calling All Admins (September 2003)
We're looking for the remotest administration: send us your latitude and longitude, the latitude and longitude of the furthest-away Linux system you have admin responsibility for, and the distance between them. The person with the furthest distance will win a lifetime subscription to *LJ* and also fun Linux schwag.

Beer Trivia (July 2003)
How many *Linux Journal* covers pictured beer on them? One winner, who will be enrolled in a Beer-of-the-Month club for one year, will be randomly selected from all correct submissions. Beer prize available to residents of the continental USA only and must be 21 years of age or older (all are welcome to enter the contest, if a winner who does not fit these requirements is selected, another prize will be presented).

LINUX JOURNAL

For more information visit http://www.linuxjournal.com/contest

Floppies for the New Millennium

Keep crypto keys, scripts and frequently used documents ready for travel on an inexpensive, rugged USB device. BY RICK MOEN

Late in 2002, my wife bought me an Easy Disk USB flash memory device built around an NAND-type flash chip with an ARM7 controller. It's a cute little plastic thing about the size and shape of a stubby cigar or a wide pen, and I routinely keep it in my pocket for convenience. This particular one has a 32MB capacity and is sold at your local PC clone shop for about $10–$20 US, but various capacities are available, ascending by powers of two all the way up to 2GB for about $600. The sweet spot on today's market is probably 256MB for about $80–$90 US; though offerings and prices are in flux as manufacturers continue to expand the market (Table 1). The smaller units seem to be vanishing as stocks run out, and the 2GB ones are just now hitting the market.

Table 1. Typical Street Pricing

Capacity	Price (US)	Price/MB
8MB	$28	$3.50
16MB	$21	$1.31
32MB	$30	$0.93
64MB	$40	$0.63
128MB	$58	$0.45
256MB	$91	$0.36
512MB	$170	$0.33
1GB	$321	$0.31
2GB	$600	$0.29

For some computer users, the gadget that finally made USB part of daily computing was a digital camera or scanner, or a USB mouse or keyboard. For me, it was this unassuming little widget. Why? Because it solves the same problems we used to solve with floppy disks, updated to modern performance and capacity standards.

Floppies themselves are obsolete on account of low capacity, speed and reliability, but the need exists more than ever for impromptu file transport between machines, especially for those of us who travel about with laptops. Ideally, everyone would have compatible 802.11a, b or g wireless or infrared networking or would be able to plug in to an available Ethernet hub, but that won't be reliably true in the near term. Even a crossover Ethernet cable, foolproof and compact as it is on the hardware level, requires software cooperation on both ends, which often doesn't happen.

Figure 1. Two USB flash drives. Left: my 32MB Easy Disk. Right: my wife's 128MB Soyo Pen Drive Pro. The Easy Disk's attachment point is on its dustcap, which thus remains on your key chain, belt or wherever, while the business end is in use. Soyo (like most flash drives, unfortunately) flubs this by putting its attachment point on the wrong end.

Iomega Zip disks are lovely for capacities up to 100MB, but most people don't have the drives. CD-R/CD-RW drives are more of an archival medium than they are casual disk storage, because one must assemble and burn session data to create them. Floppies are, by modern standards, too slow, too fallible and too tiny. So, there's been a functionality gap into which USB flash memory drives step nicely. They're fast, spacious, nonvolatile, durable, compact, cheap and compatible with all

recent PCs and Macs, regardless of operating system.

When USB first appeared on PCs, physically connecting devices like Easy Disks entailed the serious inconvenience of reaching around to the system unit's back panel to find the USB ports. You still will sometimes encounter this situation, especially on PCs whose USB ports have gone unused. People who use such ports regularly tend to attach USB hubs to them for ease of access, such as the hubs increasingly built in to current production monitors. Also, many of the newer workstation case designs move their USB ports to the front panel.

Physical Characteristics and Wear

The storage medium inside the hard plastic shell is NAND-type flash memory, which is neither fish nor fowl. It's not volatile like the classic (but exotic) electronic disk drives. Material written to a flash disk will stay good for a decade or more, with no need for AC power or batteries. It's not fragile like a hard disk, nor are there moving parts as in hard, floppy and Zip disks. Its write operations, at about 1MB/s, are much slower than those of CD-R or even CD-RW drives and are poky compared to a hard drive's. Read operations, on the other hand, are about five times faster and don't wear the device.

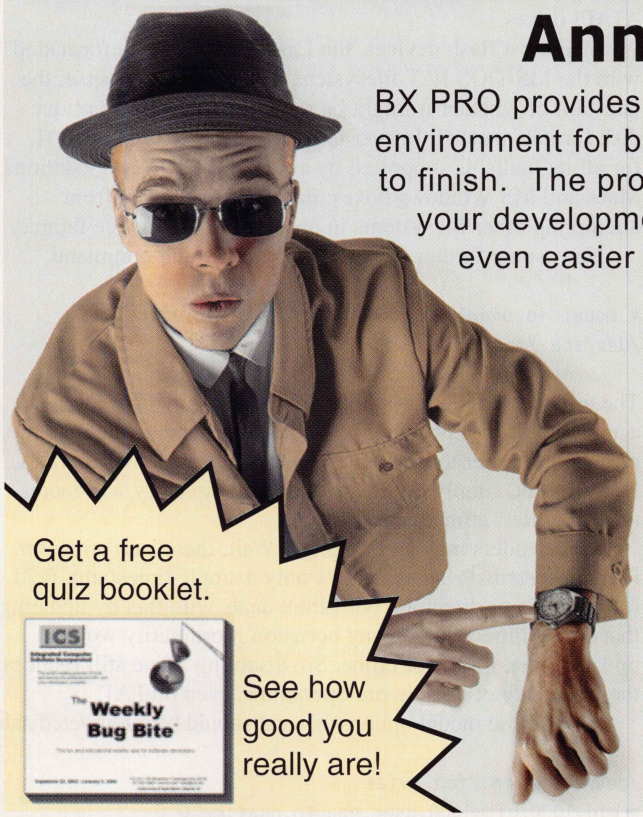
The literature on NAND flash disks suggests that they wear out after about 10,000 erase/write cycles, which just might sneak up on you, given that you can't hear any signs of distress. Any write operation requires that the onboard controller chip first zero out a fairly large data block, typically 8 or 16KB. Eventual block failure will occur from fatigue after too many erase/write cycles, at best requiring that the controller chip's hardware-level ECC functions mark as bad that entire block, and at worst causing device failure, if key filesystem information was stored there.

The point is that, although the design encourages you to treat flash disks as random-access devices, their wear characteristics are more like those of sequential media, such as magnetic tape. Accordingly, when using flash disks as Linux mass storage, you should take measures at the software level to limit wear.

In addition to the USB form factor, you'll also find NAND-type flash memory in PC Card (PCMCIA) devices—plus in a half-dozen or so closely related physical formats commonly used for data storage in digital cameras, PDAs, cellular phones and the like: CF (CompactFlash), MMC (MultiMedia Card), SD (Secure Digital), SmartMedia, Memory Stic, XD-Picture, Microdrive and Memory Gate. Most and perhaps all of those latter flash types omit the logic circuitry that enables USB flash disks to self-monitor for ECC purposes, support boot code and so on, being closer to simple flash storage with a standard access port. Most mentions of flash devices you'll encounter will turn out to concern CF-type media or similar; take care not to confuse these with USB flash drives.

Given a 64MB USB flash disk, one should be able to put entire Linux mini-distributions, such as the LNX-BBC, on

20
Got ~~30~~ Minutes to Build a User Interface?
Announcing BX PRO 6.1



BX PRO provides a superior object-oriented development environment for building powerful user interfaces from start to finish. The proven functionality of BX PRO will accelerate your development and version 6.1 is more powerful and even easier to use!

With BX PRO you get:
- ▶ WYSIWYG Visual Development Environment
- ▶ Large Project Support
- ▶ Optimized Environment for C/C++ Development
- ▶ Motif 1.2/2.1 Support

Get a free 30-day trial right now:
www.ics.com
or call
800.704.6824

ICS
Integrated Computer Solutions Incorporated

LINUX JOURNAL Editors' Choice

Copyright 2002 Integrated Computer Solutions, Inc. All rights reserved, all trademarks are the property of their respective companies.

them. If your machines have BIOS support for booting from USB, it might be worthwhile to experiment.

Configuring USB on Linux

Some of us old-fogy x86 Linux users have been slow to warm to USB. In the 2.2 kernel days, we initially heard of people needing to solve driver problems to benefit their fancy digital cameras and thought "better them than us". Linux USB support was first pioneered by Inaky Perez Gonzalez and then rewritten by Linus Torvalds, Greg Kroah-Hartman and others, and was initially rough going.

Then, USB scanners, printers, mice, video cameras, ISDN and xDSL devices, modems, floppy drives, loudspeakers, joysticks, digital tablets, MP3 players, wireless devices, PDAs, sundry mass-storage devices and just about everything else hit the market—USB started looking a whole lot less like the Useless Serial Bus we used to joke about. Fortunately, the kernel's driver support caught up in the interim, but its architecture may be a bit unfamiliar to many. If you're lucky, your distribution will have made the messy details fade to background, but in case it hasn't, read on.

Linux's USB support starts with the kernel needing to recognise your motherboard's USB chipset, which will be a UHCI (Intel) or OHCI-class (Compaq and others) device, requiring the usb-uhci or usb-ohci kernel driver, respectively. (Both also need the usbcore driver.) If `lspci -v` returns USB information that includes `I/O ports at`, then you have a UHCI controller. If the returned USB-controller text includes `Memory at`, then it's OHCI.

When you're done tweaking module loading (if necessary), the output of `lsmod` should include all three required drivers. For example, my laptop machine lists:

```
Module                  Size  Used by    Not tainted
usb-uhci               20676  0  (unused)
usb-storage            97120  1
usbcore                48000  1  [usb-uhci usb-storage]
```

The usb-storage driver is a translator that lets random-access-type USB devices be addressed using SCSI block-device names—we'll return to that later. Without it, you'd need some other intermediary to access files from the USB layers, such as the gPhoto2 application.

If you're running a 2.3.38 or later kernel (and you should really upgrade to 2.4.x or later, at this point), you also should add the following line to /etc/fstab to enable USB device tracking:

```
none  /proc/bus/usb  usbdevfs  defaults  0  0
```

After this, type `mount -a`. Now, you're all done except for mounting the actual mass-storage device. The above step does not mount the device—usbdevfs is strictly an abstract support filesystem similar to /proc, used by the USB subsystem.

All of the above USB-configuration details should be taken care of automatically by modern Linux installer programs, but it is covered here in case it's not. If you're reasonably lucky, all you need to do is plug the drive in to a USB port or hub and mount it.

Mounting and Managing the Flash Drive

I created a mountpoint directory of /mnt/fob from which to hang the flash drive. That name owes to the drive being about the same size as a key fob or an old pocket watch (fob watch), lurking in one's trouser pockets in roughly the same fashion.

That was the easy part. Next was a wild ride trying to chase down how to mount it properly. My browsing of on-line materials suggested that USB flash drives, as with Iomega Zip disks, would be treated as SCSI. A glance at /proc/scsi/scsi after plugging in the Easy Disk clarifies that it's recognised as the first SCSI device and moreover that it's a rebranded DiskOnKey unit OEMed from M-Systems, Inc. Intuition therefore suggested mounting /dev/sda1 (on an otherwise IDE-based system), but intuition turned out to be dead wrong in the Easy Disk's case. Making the attempt resulted in this puzzling error:

```
guido:~# mount -t vfat /dev/sda1 /mnt/fob
mount: block device /dev/sda1 is write-protected,
mounting read-only
mount: /dev/sda1 is not a valid block device
```

Some other USB flash drives reportedly do mount precisely in that way, whereas I found out that the Easy Disk mounts as /dev/sda, not /dev/sda1 or any other partition number—that is, it treats the drive as a device lacking the ability to house a partition table. The explanation of this curiosity turns out to lie in the ATAPI Removable Media Device (ARMD) BIOS Specification, laid down by Compaq and Phoenix in 1997 to generalise how a floppy-like drive should behave on the ATA bus in anticipation of traditional floppy drives going the way of the dodo. ARMD devices are basically big floppy disks, which per the spec have no partition table. Thus, the Easy Disk happens to be an ARMD (floppy-like) device. Apparently, many other USB flash disks are, by contrast, treated as regular ATAPI drives.

Like most flash devices, the Easy Disk came preformatted with the MS-DOS FAT filesystem. That raises, of course, the question of whether it might be rewritten to use less antique disk formats instead. I never tested that, partly because FAT, for all its faults, is supported by almost all UNIX workstations, Macs and MS Windows boxes. Partly, I shied away from attempting other filesystems in order to reduce device fatigue.

Now, some further elaboration of the mount command:

```
# mount -o uid=1000,gid=1000,noatime -t vfat \
/dev/sda /mnt/fob/
```

The uid and gid mount options are those of my own login account to set file ownerships. More significant is the noatime, provided in order to eliminate unnecessary erase/write cycles. Don't forget, simply doing an `ls` on any directory will increment the files' atime (access time).

Alert readers may be thinking: "Wait, there's no atime on FAT filesystems!" True. There's only a single timestamp field for each file or directory, but Linux deals with this by updating that single time value on any occasion it ordinarily would update atime, mtime or ctime. So, disabling atime still reduces the frequency of erase/write operations, even on FAT.

All of those mount options can and should be put in /etc/fstab:

```
/dev/sda  /mnt/fob  vfat
➥uid=1000,gid=1000,user,noauto,noatime 0 0
```

One further oddity—no matter what you do, the flash disk always mounts read-only:

```
mount: block device /dev/sda is write-protected,
mounting read-only
```

This happens even if you specify rw among the mount options. However, you subsequently can enable write access after mounting the flash disk, by remounting with the rw option:

```
# mount -o rw,remount /mnt/fob
```

Exactly why /bin/mount insists that the flash disk is write-protected and must be mounted read-only is a genuine mystery. Although some flash disks' plastic casings reportedly sport write-protect switches, the Easy Disk's doesn't. My best guess is that mount is heeding a request from the Easy Disk's built-in controller chip, intended to minimise accidental device fatigue. And it works. As a side benefit, the read-only default seems to render harmless your unplugging of the device when, inevitably, you forget to unmount it first—making it truly a hot-plug device.

The Hard Drive in My Pocket

Having coped with the hurdles and minor oddnesses of getting Linux support configured for the flash drive—or pen drive, as they are sometimes called—what strikes one most about these devices is how they fade to background. You simply rely on them and take them for granted, which is the mark of any truly successful technology. Documents and applications you use frequently, GnuPG and other crypto keys and files you need to transport among computers, regardless of operating system, are stored on the flash disk and dropped into your pocket. You don't have to worry about magnetic fields, mechanical shock, spontaneous bit-rot or anything else. It simply plugs in to a free port and works. There's nothing else quite like it.

Rick Moen is a sysadmin, writer and IT guy in the San Francisco Bay area where he has been a longtime member of its Linux community for which he runs an on-line calendar of upcoming events, BALE (Bay Area Linux Events, linuxmafia.com/bale).

RESOURCES

Article about using Linux on a flash drive that registers as /dev/sda1 instead of /dev/sda: www.gctglobal.com/Download/3rd_LED/PalmKey/palmkey.html.

ATAPI Removable Media Device (ARMD) BIOS Specification, formerly the ATAPI Removable Drive (ARMD) Specification: www.phoenix.com/resources/specs-atapi.pdf.

Linux also supports CF devices and can boot from them. See the Memory Technology Device (MTD) Subsystem for Linux Site: www.linux-mtd.infradead.org.

Anti-Virus Anti-Spam
Anti-Suck

Customized Network Protection

We do the ground work for you, to find the best anti-virus and spam fighting solutions available and we support each and every product we endorse.

Excellent competitive discounts available to current RAV users: we don't forget our customers.

Anti-Products (spam and virus busters):

F-PROT antivirus
F-PROT
- Highest Level of Anti-Virus Research
- Unobtrusive, affordable Windows antivirus protection
- Sendmail LibMilter, Qmail and Postfix anti-virus protection
- System level and daemon based virus scanning for Linux, xBSD, AIX and Solaris

bitdefender secure your every bit
BitDefender
- Centrally managed Windows desktop anti-virus protection
- Linux Sendmail, Qmail, Postfix and SMTP Relay anti-virus protection
- Protect your entire environment with versatile Corporate License Pack

CanIt
- Stops up to 98% + of spam
- Unlimited administrative and/or end-user control
- Ability to filter in-bound and out-bound e-mail
- Customizable, multi-level spam protection

SafeInternetEmail SERVER
SafeInternetEmail
- Integrated anti-virus/anti-spam/pattern scan engine
- Per-Domain and Per-User controls
- Dynamically learns your email patterns
- Military proven technology

20% Off all online sales, use discount code 00AXE. Applies to online sales only. Call 1.877.302.2027 or visit:

www.raeinternet.com

raeinternet

Tastes great, more filling.

Start your subscription today:
http://www.linuxjournal.com/

Reviewing the Basics

Whether you're new to Linux or looking to brush up on some of the basics, the *LJ* Web site has articles to help. **BY HEATHER MEAD**

A few months ago, a friend of mine who knows I work for *Linux Journal* asked if I could help him install a Linux distribution on his home computer. He's a part-time student and didn't want a repeat of a situation from last semester: having rebooted his system after it crashed, he discovered the document file he'd been writing was corrupted. This wasn't the first time he'd "lost" work, and he wanted to know if there was another OS option. We switched him to the latest Red Hat, which uses the journaling ext3 filesystem, and so far things have run smoothly.

I thought of my friend recently when reading Bruce Byfield's Web article on "Breaking the Word Processor Curve" (www.linuxjournal.com/article/7120). Bruce builds a good case for OpenOffice.org's Writer being the best word processor around, because it takes a completely different approach to style, structured texts and long documents. Instead of using menus to access a limited number of style options, OOo Writer "offers a floating palette called the Stylist. Repositionable anywhere on the screen, the Stylist not only makes the application of styles more convenient, but it makes the editing and creation of styles a single right-click away." In addition, many features that don't work in Microsoft Word do work in OOo Writer, such as master documents and automatic numbering. In this case, although my friend was looking for an alternative, he may have ended up with the better choice.

If you're already familiar with Linux or other UNIX systems but want some refresher articles, start with "Using CFS, the Cryptographic Filesystem" (www.linuxjournal.com/article/6381). Author Jerry Sweet explains how CFS encrypts files in a directory and then lets you decrypt them to plain text for a specified amount of time with a local loopback NFS mount. When you're done, you check the files back in, and they are once again encrypted. If you're interested in trying CFS, Jerry's article also provides scripts and tips.

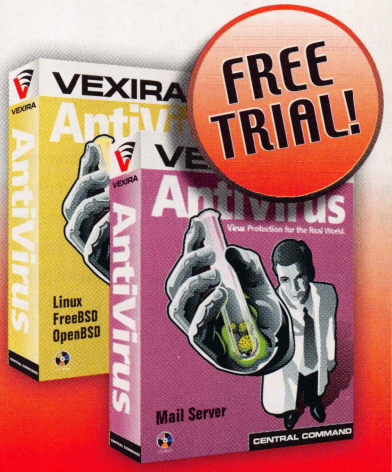
Two more good refresher articles are "Overview of Linux Printing Systems" (www.linuxjournal.com/article/6729) and "Linkers and Loaders" (www.linuxjournal.com/article/6463). The first article discusses the basics of UNIX printing systems and how they revolve "almost entirely around the PostScript page description language". Author Stephane Peter then shows how these basics apply to Linux and covers the BSD LPD printing system, the LPRng printing system and CUPS. "Linkers and Loaders", on the other hand, is an overview of how compilers, linkers and loaders work. Author Sandeep Grover explains three different types of object files and the difference between working with static and shared libraries. If you're looking to get into development or administration, these two articles might be a good place to start your research.

Much of our lives are spent learning new things and teaching others, and as far as Linux goes, the *Linux Journal* Web site is here to help you. If you'd like to write an article that covers a basic or teaches us something new, drop us a line at ljeditor@ssc.com.

Heather Mead is senior editor of *Linux Journal*.

Best of Breed

Virus Protection for Mail Servers



Vexira™ Antivirus for Mail servers is an industry leading, highly scalable virus scanning solution for email servers. Contact Central Command® today to find out why Vexira Antivirus is known as the "best of breed" in virus protection.

Supported OS:
- Linux
- FreeBSD
- OpenBSD

Email Servers:
- Sendmail
- Postfix
- Qmail
- Exim

Central Command, Inc.
phone: 330.723.2062 • fax: 330.722.6517
www.centralcommand.com

© 2003 Central Command, Inc. All rights reserved. Vexira, Vexira logo, and Central Command are trademarks or registered trademarks of Central Command, Inc. All other trademarks are property of there respective owner(s).

Setting the System Time over the Network

I have installed ntpd on Red Hat 8, and it runs but does not appear to do anything. The log file indicates it is starting up, but after a few minutes, it displays the following message:

```
kernel time discipline status change 41
```

After that, nothing else is written to the log file. What does the kernel time discipline message mean? How can I tell if ntp actually is updating my system clock? If it isn't doing updates, how can I get it to work properly?

--

Lance Hill, lancemhillnospam@hotmail.com

To check if ntpd, the Network Time Protocol dæmon, is working correctly, type `ntptrace localhost` *to see where your system is getting its time and where its upstream time server is getting its time. To be polite, don't set up ntpd to use a heavily loaded public time server—your ISP or local Linux user group should be able to give you access to a server that is close to you.*

--

Don Marti, dmarti@ssc.com

Here are a few things to check. First, ntpd will not synchronize time differences greater than one hour, so use ntpdate first to get your clock in range. Second, make sure you have not configured ntpd to use multicast/broadcast client functionality unless you have a time server on your local network. Third, if you are running a recent version of ntp make sure you have a restrict parameter in your configuration file for each time server you use, as newer versions are much more security-aware. Finally, use the ntpq tool and the assoc command to verify that ntpd is running properly.

--

Chad Robinson, crobinson@rfgonline.com

Exposing an Internal Site with Squid

My wife's computer is inside a LAN and accesses the Internet through a proxy (no other connections are allowed, only WWW). As we are working on a common project, I need her to connect to some of our internal sites that are accessible only through another proxy. The situation is this: my wife's computer→proxy1→Internet→proxy2→internal sites. In that scheme, I have full control on proxy2 (a Debian box running squid). Any ideas about how to do this?

--

Mauro A. Cremonini, mac@foodsci.unibo.it

If you are running the Squid proxy, cache_peer and other settings allow you to daisy-chain proxy servers together. You need to establish a cache_host entry that defines the internal sites as described above and a dstdomain ACL entry for the internal sites that forces them to be accessed through proxy2. Squid is a powerful proxy server, and you can slice and dice your traffic management a number of different ways. For more tips and tricks, see the Squid FAQ at www.squid-cache.org/Doc/FAQ/FAQ.html.

--

Chad Robinson, crobinson@rfgonline.com

*If you aren't running a Web server on proxy2's port 80, you can run rinetd to forward connections from it to a system on your internal network (*www.boutell.com/rinetd*). For security, though, the best thing would be to ask your wife's network administrator to open up their firewall for outgoing SSH. Then, give her a shell account on proxy2 so that she can tunnel in with the following:*

```
ssh -N -L 8080:internal:80 proxy2
```

Here's an explanation of the ssh command-line options above: -N, don't start a shell; -L, forward a local port; 8080, port on which to accept connections on localhost; internal, host to which to forward connections; 80, port to which to forward connections; and proxy2, host to which to make the SSH connection.

Then, she simply can point her browser at localhost:8080 and get to your internal server, and you don't have to open up anything except incoming SSH to your proxy.

--

Don Marti, dmarti@ssc.com

Printing from an Application

I'm trying to port some Windows applications to Linux (Red Hat 7.3 and 9), and I need to find a replacement method for my current printing mechanism. The current printing process goes like this: 1) The application opens LPT8: for write as a file. LPT8: is a Windows device, not a real physical port. Data (plain ASCII text) is written to the device. When done, the device is closed. 2) LPT8: is captured by a Novell Netware capture program so that anything written to LPT8: winds up in a Novell Netware printer queue. 3) The Novell Netware server sends the data to a barcode printer. The printer is attached to the network using one of those little AXIS print servers that can handle multiple protocols including Novell NPRINTER and lpd. I want to know if there is any way to keep the basic operation of the program the same: open a file/device, write to it, close it. Everything else that needs to happen occurs in the background. I can see doing this using calls to external programs: create a temp file with the data, then pass the file to a printing program like lpr. This is a bit more complicated but will work if the first method doesn't. Or, one could create a temp file and put a watch on the temp directory to grab any files there. The user expects near instantaneous response, however. The Novell Netware server is going to be replaced by a Linux box eventually. The Windows machines will use Samba for their printers. I haven't decided on which printing system to use yet, probably CUPS. The programs are written in Delphi/Kylix Object Pascal. Any hints on where I should start looking?

--

Gus Wirth, gwirth@sciti.com

This can be easy or difficult depending on what you are willing to do. The solution you referenced above that involves the creation of a temporary file actually is the easiest solution. You can use the tmpnam() function to obtain a unique filename to ensure your process won't conflict with any others. This concept keeps you within the open-file, write-file, close-file work flow you described above, and you shouldn't require an external program to do this work. If your program was structured such that it always created a temporary file and called out to a configurable command, it would remain portable. On a Windows box one could use a copy command to dump the file to LPT8:, and on

a Linux box one could use the lpr command. This would make your program more flexible.

--

Chad Robinson, crobinson@rfgonline.com

Running a Laptop without ACPI

I installed Red Hat 9 on a Dell Inspiron 1100 laptop. It appears that no power-management support was installed (this is an ACPI machine, not APM). A few Web sites have reported success with this system, except for ACPI. I tend to stick with Red Hat standard releases because I rely on my laptop for everyday use (I use it 6–8 hours a day for teaching). I usually don't experiment on my working system. What is the impact of not including ACPI support? Do I run the risk of damaging the CPU without it, or can I live without it (at least for a while)? Can you point me to a how-to about ACPI and Linux?

--

Erik Hemdal, ehemdal@townisp.com

Running with ACPI enabled will not harm your system unless you are overclocking or doing some other activity that relies on the alerts ACPI would provide. In a laptop environment this is extremely unlikely. However, you may not see the same battery lifetimes without ACPI, because that is the same module used to identify whether the system is running on AC or battery, to scale the speed of the CPU and so forth. For more information on ACPI, consult the ACPI-HOWTO in the Linux Documentation Project (LDP): www.ibiblio.org/pub/Linux/docs/HOWTO/other-formats/html_single/ACPI-HOWTO.html.

--

Chad Robinson, crobinson@rfgonline.com

Try www.linux-laptop.net, *a Web site where people list problems they experience while installing Linux on laptops.*

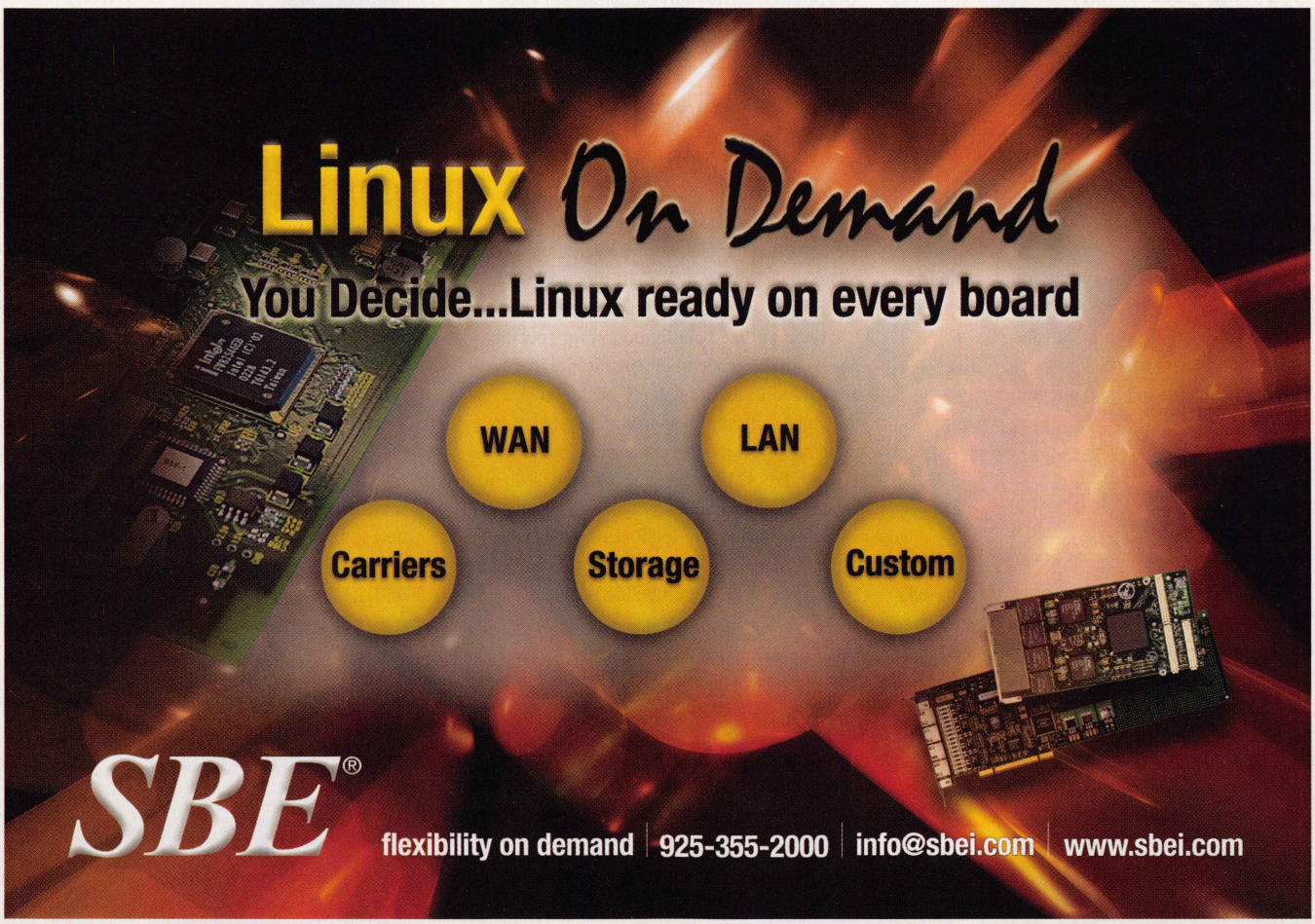
--

Usman S. Ansari, uansari@yahoo.com

Many on-line help resources are available on the *Linux Journal* Web pages. Sunsite mirror sites, FAQs and HOWTOs can all be found at **www.linuxjournal.com**.

Answers published in Best of Technical Support are provided by a team of Linux experts. If you would like to submit a question for consideration for use in this column, please fill out the Web form at **www.linuxjournal.com/lj-issues/techsup.html** or send e-mail with the subject line "BTS" to bts@ssc.com.

Please be sure to include your distribution, kernel version, any details that seem relevant and a full description of the problem.



Atmel Embedded Linux Port

Atmel Corporation announced the availability of a Linux port for its AT91RM9200 microcontroller and its ARM9 core-based ASICs. The port consists of the 2.4.19 kernel, the ARM9 Linux adaptation maintained by the ARM Linux support team, a kernel patch for the advanced interrupt controller and system timer, and a set of AT91RM9200-specific drivers. The port also includes a minimal RAM disk that contains the most frequently used commands. In addition, a development kit is available that contains the port, the U-Boot loader utility and the GNU toolchain in source and binary forms.

CONTACT Atmel Corporation, 2325 Orchard Parkway, San Jose, California 95131, 408-441-0311, www.atmel.com.

Zend Studio 3.0

The Zend Studio 3.0 IDE now is available for Web developers building mission-critical applications in PHP, JavaScript and HTML. Zend Studio 3.0 supports PHP 5 and offers increased performance speeds and improved project management tools. Version 3.0 also has new code debugging tools, including a remote debugger that provides developers with a debug URL so they can debug directly from a browser. In addition, version 3.0 includes a Code Profiler, which tracks down pieces of the application that are slowing down the project, and a Code Analyzer, which analyzes static source code to reconcile problems. Other new features include improved XML integration, advanced code completion, customizable key maps, full version management and FTP integration.

CONTACT Zend Technologies Ltd., PO Box 3619, Ramat Gan, Israel, 52136, 877-936-3747, www.zend.com.

QVISION 2.0

QVISION version 2.0 is a real-time, view-based network monitoring tool that provides visibility into all the activities on networks, connected systems and applications. QVISION resides on the network at the Internet connection and inside the network, monitoring traffic flows and detecting unnatural network patterns. New features for QVISION 2.0 include a behavior view, a non-signature-based method of exposing network, policy and security issues based on equations linking event properties in real time; intelligent alerting, triggers set off by specific predetermined behaviors; and selective application content capture, a tool that allows users to configure the desired level of content capture.

CONTACT Q1 Labs, Inc., 15 Piedmont Center, Suite 1040, Atlanta, Georgia 30305, 877-471-5227, www.q1labs.com.

Raptor 74GB SATA Drives

Western Digital released a new version of its serial ATA (SATA) 10,000RPM WD Raptor hard drives. The new Raptor drives offer 74GB of storage capacity, Ultra/150 Command Queuing (CQ) technology and Rotary Accelerometer Feed Forward (RAFF) technology. Raptor drives are designed for enterprise-level specifications, with 1.2 million hours of mean time between failures, 10,000RPM, an average of 4.5ms seek time and a five-year warranty. Ultra/150 CQ technology optimizes the sequence of data transfers between the host and hard drive for increased transfer efficiency. RAFF technology maintains hard drive performance in high-vibration environments. In addition, the Raptor 74GB drive utilizes fluid dynamic bearings to reduce noise and increase reliability.

CONTACT Western Digital, 20511 Lake Forest Drive, Lake Forest, California 92630, 949-672-7000, www.westerndigital.com.

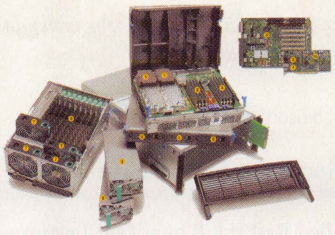
Metrowerks Embedded BSP for Motorola

A new board support package (BSP) from Metrowerks is designed to help developers create embedded applications powered by Motorola's PowerQUICC III processor. The Metrowerks BSP is tested and configured for use within a fixed configuration for specific hardware reference platforms. It includes an embedded Linux kernel, a GNU debugger and toolchain, a runtime environment and a complete set of device drivers and libraries for PowerQUICC III processors. Aimed at the networking and communications markets, the BSP works with Metrowerks' Platform Creation Suite, a management tool that helps developers customize Linux for embedded devices.

CONTACT Metrowerks, 7700 West Parmer Lane, Austin, Texas 78729, 512-996-5416, www.metrowerks.com, sales@metrowerks.com.

Glacier Itanium 2 Systems

Aspen Systems announced the release of two new systems based on the Itanium 2 processor and targeted at the high-performance computing (HPC) market. The Aspen



Systems Glacier Dual Itanium 2 system has a 1.4GHz processor and 1.5MB of L3 cache. Based on the SR870BH2 Intel server platform, this system is available in a 2U rackmountable form factor. The second system is based on the Low Voltage Itanium 2 processor and comes in 2U and 4U rackmountable form factors. Its processor has a 1.0GHz core clock frequency and 1.5MB of L3 cache. Both systems have low power consumption, reducing the cost of scalability.

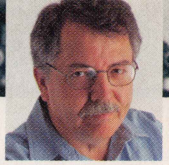
CONTACT Aspen Systems, Inc., 3900 Youngfield Street, Wheat Ridge, Colorado 80033, www.aspsys.com.

Linux for Everyone 3.0 with Cognitio

CPUBuilders released a new control panel, Cognitio, with its Linux for Everyone v 3.0 systems, which are available through Sam's Club and the CPUBuilders Web site. The Cognitio control panel is designed to handle ease-of-use issues relating to Linux-based hardware and software. It allows users to add plug-and-play devices, CD writers, hard drives, scanners and USB storage. Also new in v 3.0 Linux for Everyone systems are USB 2.0, faster processors, DDR memory on all models and USB 6-in-1 card readers. In addition, 3.0 systems feature one-click software updates, automatic monitoring tools for network and Internet connections, automatic network file-sharing access points and documentation for new Linux users.

CONTACT CPUBuilders, 316-315-0300, www.cpubuilders.com, support@cpubuilders.com.

Please send information about releases of Linux-related products to Heather Mead at newproducts@ssc.com or New Products c/o *Linux Journal*, PO Box 55549, Seattle, WA 98155-0549, or fax to 206-782-7191. Submissions are edited for length and content.



Free Business

Quality and service—what kind of a software business is that? **BY DOC SEARLS**

I'm writing this column from a hospital, specifically, its Critical Care Unit (CCU). They used to call these intensive care units, or ICUs. Before unit became current, they were wards. All the name changing causes confusion. I wonder if it's enough to cost lives.

It's conceivable. Two doors down, they had a Code Blue this morning. Machines were wheeled in. Staffers hurried around, carrying clear tubes and bags of fluid. A chaplain paced in the hall. Groups of family members cried on each other's shoulders. They deal with life or death in here. A hospice worker a few minutes ago told me "a lot of the people in there are wheeled out feet first."

Fortunately for me, I'm not a patient here. Unfortunately for my mother, she is, thanks to complications following a gallstone removal procedure. So I've been hanging around as a visitor for most of the past week. Every once in a while, I go over to the counter by one of the nursing stations, where a Microsoft Windows 2000 Professional logo floats on a black screen above a black keyboard and a mouse. When I move the mouse, the screen comes to life, offering me a choice between Staff Login and Internet Explorer. After clicking on Explorer, I surf the Net until it crashes on some site, bringing the rest of the machine down with it.

I asked a nurse yesterday if this kind of crash also happens when the systems are running the hospital's own applications, which mostly are used to access patient records and recent lab results. "Those things are down all the time", she said. "We have all kinds of problems with them." After hearing her response, I wondered, "what are costs—in time, money and human lives—to running hospital computers on a crashy operating system?"

Three weeks now have passed between the last sentence and this one. Several days after I wrote everything above, my mother's condition improved to the point where she could be moved out of the CCU. She had fought hard against pain so extreme that she needed frequent doses of morphine, which were shot into one of the many stretches of tubing that ran in and out of her frail, 90-year-old body. But she was tough and willful; her mind still was strong and amazingly lucid, especially considering all the chemical insults it endured.

"Are you gonna make it, Mom?" I asked. She said yes and told me to head back home to California. That night, best they can figure, she had a stroke. Two days later her cardiologist came by, surprised to find her there. He hadn't been told she was in the hospital and therefore hadn't been consulted about her medications, which he carefully had prescribed and monitored over the last ten years, ever since she had been diagnosed with congestive heart failure. The next day, she died.

It's easy to second-guess everything at times like this. Any one of us—doctors, nurses, relatives—could have made sure to involve the cardiologist. And, as many doctors and nurses have told us (some of them members of our own family), Mom's odds were long at best from the moment the original procedure went poorly, which often happens with elderly patients. She knew the risks going in. She also lived a full and happy life. Before she slipped out of consciousness, she told my sister that she had no unfinished business and was ready to go. Her last word was love. She passed peacefully and without pain.

So I have no complaints about her treatment, which was excellent. I also know that every botched procedure, every course of treatment that fails to reach the ideal outcome—and, let's admit it, there are many—is a learning experience. One thing I would like hospitals to learn is how to get along without crashy software.

My mother died in late August 2003, during a time when the Blaster/Lovsan worm was infecting and spreading among countless Windows computers. A few days later, on "The Linux Show" (where I'm a weekly regular), Bruce Sterling told us there is no overestimating the level of hatred for Microsoft that the latest round of viruses is causing, especially outside the US. He recently had returned from Europe, where he found large cadres of locals determined to move beyond dependency on Microsoft. Bruce also wondered out loud whether there would be any kind of software business at all when "open-source nirvana" finally was achieved. Here's how he put the situation in a *Wired* piece titled "Freedom's Dark Side: The iron fist, the invisible hand, and the battle for the soul of open source" (www.wired.com/wired/archive/11.09/view.html?pg=4):

> Logically—indeed, free-software geeks are the most logical hippies in the whole wide world—the revolution is at hand. Why should anybody pay for software? What do you get for your money besides shrink-wrap licenses, potential lawsuits, DRM cuffs around both wrists, and a cloud of viruses? "Property relations" are blocking social and technical progress. Secure computing and digital rights management are coercive regimes that would make George Orwell blush. The free market is a tissue of political fiction as brittle as an Eastern European regime. With open source code on tap, the software trade will collapse under its own weight....
>
> From Redmond or Cupertino, it may seem decadent that an elitist scrum of hairy cyber-Euristas would style itself a cultural avant-garde. But consider the track record. In 1989, European civil society tore the guts out of Communism. When the Europeans won the Cold War, they got a bonus the US never did: zillions of eager former Commies. For these New Europe types, open source is a great, glittering step up from the vile product they're used to: *broken-open* source software, as in pirated CDs sold off blankets at flea markets.
>
> The denizens of Open Cultures want their connected collectivism to liberate the world from regulations, markets, and

> intellectual property. But what if victory only clears the way for corruption of their beloved culture?...I have to wonder what dark passions and ancient evils have been held in check by the grim totalitarianism of the profit motive. We may yet find out.

Clearly, Microsoft's operating system and the PC productivity software business are threatened, or Microsoft wouldn't repeat "Linux" and "threat" so often in the same sentences. But it's a huge mistake to assume they are the only software businesses worthy of the label. What about that hospital software? What about the database, corporate accounting, customer resource management and transactional computing stuff that companies all over the world use to run their businesses? Do these all have to run on Windows? Do even a few? Look at Oracle, which abandoned its formerly nonpartisan orientation to operating systems and embraced Linux with a whole heart and a fattening wallet. Last I looked, Oracle still was growing like a weed and making billions of dollars on margins they've kept north of 75%. Look at Amazon and Orbitz, both of which run everything they can on Linux.

These companies, of course, are the obvious ones. How about the countless small- and medium-sized businesses and the categories they occupy? I'm talking here about stuff that stays way below the radar, because what they do isn't sexy or leading edge, even if they're the folks whose listings fatten our yellow pages.

Take, for example, the junkyard business. Last week, I was talking with my friend, Allan Harrell, in Phoenix, who has ties with body shop, body parts and junkyard businesses, even though his main job is keeping small business networks and computer systems up and running. After telling me how full his plate has become, thanks to all the viruses infecting Windows machines in his care, he mentioned FastParts, Inc. (fastpartsinc.com), the country's second largest supplier of computerized inventory systems for junkyards.

Intrigued, I got Jim Mangham and Wayne Leland on the phone to tell me about it. Jim owns the company, and Wayne is the "lead programmer systems guy". They have a staff of seven and a customer base of 750. Wayne said the company has been using Linux for six or seven years; they like being able to hack the kernel and otherwise customize systems to do what they want.

FastParts sells a service that's delivered with a turnkey system in a white box. Customers view and operate it in a terminal window on whatever system (usually Windows) they happen to be using, including dumb terminals (Wyse, Sherwood) that FastParts is happy to provide.

FastParts is a personal service business, delivering and installing the systems themselves, including the modem phone connection that lets the company dial into a machine, do diagnostics and fix problems. Customers also call in through the system to look for car parts on FastParts' central system. "Many of our customers are too far out in the sticks to get a high-speed Internet connection", Wayne says. "We have one guy on the road full-time, just installing."

How do they compete? "Reliability, pricing and support", Wayne says. "Out of all our competitors, where we get 'em is with support. Most of the other ones have modern telephone systems with all these menu choices you have to go through, and most of our customers do not like that. They want to get a real person when the phone rings. With us they get that."

To me, that's the real story of Linux and open source. It's the ability of any expert in any field that requires any computerization at all to build a business that delivers what customers want—thanks to underlying software building materials that are free.

One of these days we'll all understand what's been obvious for years to guys like Wayne Leland: that free software is the best bedrock for a healthy and free computing marketplace—regardless of whether it threatens prevailing monopolies.

Doc Searls is senior editor of *Linux Journal*.

More Signs of Desktop Inevitability

Sooner or later one of the Big Four (Dell, Gateway, HP, IBM) will market Linux to end users in a big way. In the meantime, of course, a huge opening exists for smaller companies with bigger plans. For the last year Lindows.com has been pushing hard into this space. Now comes Linare Corporation (linare.com), a three-year-old Seattle-area company. In September 2003, Linare announced Linare Linux Desktop OS, selling for $19.95 US, in advance of a planned rollout of a "high-quality, powerful, and fully loaded desktop PC for $199.95 US".

Linare's distribution is based on Red Hat and KDE and bundles OpenOffice.org, the Gaim instant messenger, the Mozilla browser and mail software, games and a lot of multimedia software, including RealPlayer.

Linare CEO Soma Sundaram said the company studied the market carefully and decided the reason desktop Linux hadn't taken off was a "lack of support from the OEM side". He also said recent rounds of viruses designed to attack Windows systems have created more demand for Linux, which he called "a virus-free operating system". So he decided "to take a different approach in terms of hardware and software. Anybody can install the Linare OS on any hardware today; but in the long run it's best to provide unique hardware as well."

The company plans to market its new products primarily to end users and support the effort with national television advertising.

—DOC SEARLS

Lindows MobilePC/ServeLinux eNote

REVIEWED BY STEVE R. HASTINGS

The Lindows MobilePC is a lightweight and inexpensive notebook PC that comes with Linux preinstalled. It received the 2003 *Linux Journal* Editors' Choice Award for Best Mobile Device. If you want a small notebook PC without Microsoft, you should consider this one.

I ordered my Lindows MobilePC from GearZoo (gearzoo.com). I had no problems with the ordering process, and the notebook arrived quickly. The notebook actually is a ServeLinux eNote, model ISNB-001. It is small and light: 10.43" × 8.66" × 0.91", 2.9 pounds. It uses a VIA C3 processor, an Ezra core running at 933MHz, a VIA chipset motherboard and a 20GB hard drive. The eNote comes standard with 256MB of RAM, but there is an empty socket for a standard SODIMM card. You can put up to a 512MB card in this socket, for a total of 768MB of RAM. The eNote has a single CardBus PCMCIA Type II socket, plus a CompactFlash Type I socket. Two USB 2.0 (high-speed) sockets and a non-powered 1394 socket are also present, but no legacy ports (serial, parallel or PS/2).

The 12.1" display is 1,024×768 pixels with 24-bit color. This display is sharp and clear, with bright colors. The built-in video adapter is a Savage 4 AGP chipset, which XFree86 supports well. A standard 15-pin analog video out socket also is included. For sound there is a VIA 82C686 chip with built-in speakers, a microphone jack and a headphone jack with volume control dial.

The keyboard has comfortable keys—it's still a notebook PC, so you don't type as quickly as you would on a full-size desktop keyboard, but it's quite usable. Most keys are right where I expect them—for example, I don't press the up Arrow key when I'm trying to press the right Shift key. The keys are quiet too, making only soft clicks as you type; it's not too noisy for use in a library.

The built-in pointing device is a trackpad that has two buttons with a rocker in between; the /dev/psaux device connects to it. X sees the rocker as scroll wheel events, not as a third mouse button. When I want to press the third mouse button, I use the standard X workaround of pressing both buttons at the same time.

The included lithium-ion battery pack is supposed to power the notebook for two hours, but it actually lasted closer to an hour and a half. Extra battery packs are available for about $90 US. The owner's manual describes an extended life battery that covers the entire bottom of the notebook, but I have not been able to locate where to buy one of these.

Not only is the notebook itself small and light, but so is the AC power adapter, thanks to a slim and lightweight transformer brick. Also important is a power cord long enough to reach inconvenient hotel-room power sockets. The adapter accepts 100–240 Volt AC power at 50–60Hz, so you can use it anywhere in the world with a simple plug adapter. For safety, a security slot can be found on the left side for attaching an antitheft cable.

The left side of the eNote has a phone jack and an Ethernet jack that connect to a CompactPCI card inside the eNote with a modem and a network interface. The modem, alas, is a controller-less Winmodem, and I have not found a Linux driver that works with it. A standard PC Card modem,

PRODUCT INFORMATION

Manufacturer:
Lindows.com, Inc.

URL:
www.lindows.com

Price:
$799 US

THE GOOD

- Compact and lightweight (2.9 pounds).
- Excellent expansion with CardBus, USB 2.0 and 1394 ports.
- Clear, sharp screen and nice keyboard.

THE BAD

- Cooling fan sometimes noisy.
- Battery life could be longer.
- No floppy or optical drive included.

however, worked well for me. The network interface is a Realtek 8139 chipset, and the Linux 8139too driver works with it just fine.

One of the reasons the eNote is so small and light is that it does not include a floppy or optical drive. You can purchase an external device and plug it in to either the USB or 1394 ports. The BIOS settings include the option to boot from a USB drive, a USB CD-ROM or a network boot, in addition to the hard drive. However, the BIOS only supports booting from USB, not from the 1394 port.

Using eNote

The first time I booted my notebook, I was greeted with a Lindows login prompt. No password was included in my documentation; I had to contact technical support at GearZoo. They immediately e-mailed the password I needed (which was enote, all in lowercase letters).

The 933MHz processor has ample speed for running a Linux desktop. When the notebook is connected to a high-speed Internet connection, Web pages display quickly.

A cooling fan is built in to the notebook. Normally it is quiet, but when the notebook is working hard, the fan speeds up and the noise increases. When the hard disk is busy, the fan becomes even louder. To reduce fan noise, I recommend upgrading to the full 768MB of RAM and disabling swapping.

The trackpad works well, but it is more convenient to get a small travel mouse and plug it in to one of the USB ports. I bought a Microsoft Notebook Optical Mouse with two buttons and a button/scroll wheel; when I plugged it in, Linux recognized it as a standard HID-compliant mouse and it worked.

Initially, I used the included Lindows 3.0 distribution of Linux, which appears to have little or no customization by Lindows for use with a notebook computer. For example, the Click-N-Run program attempted to use the Ethernet connection even when Ethernet wasn't plugged in. Click-N-Run also consumed a lot of CPU power and hammered on the hard disk, endlessly trying to check for updates. This was especially bad when I was running off the battery. If this happens to you, I suggest you kill the Click-N-Run process. The documentation about using a modem for Internet access doesn't mention that the built-in modem does not work; it is generic documentation from KDE.

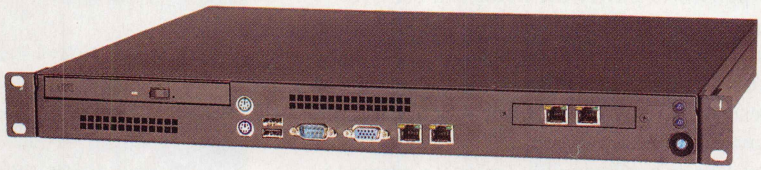
Lindows is based on Debian GNU/Linux. The Debian APT system is present, and it is possible to use the apt-get upgrade command to install a full Debian system. I upgraded my eNote to Debian's unstable branch and installed a GNOME desktop. You can read the full story of this upgrade on the *Linux Journal* Web site; see Resources for the URL.

Conclusion

The eNote is a winner. Its built-in hardware has most of the features you need, and with the 1394 and USB ports and the CardBus PCMCIA, you can plug in anything else. It doesn't have the fastest CPU available, nor does it have an included DVD or CD drive, so it's not for everyone. But if you want to be able to take notes, surf the Web or SSH in to a server, this is an ideal notebook.

Steve R. Hastings first used UNIX on actual paper teletypes. He enjoys bicycling with his wife, listening to music, petting his cat and making his Linux computers do new things.

Ultra Dense, Powerful, Reliable...
Datacenter Management Simplified!
15" Deep, 2-Xeon/Opteron or P4 (w/RAID) options



Customized Solutions for...
Linux, BSD, W2K

High Performance Networking Solutions
- Data Center Management
- Application Clustering
- Network and Storage Engines

Rackmount Server Products
- **1U Starting at $499**: C3-1GHz, LAN, 256MB, 20GB IDE
- 2U with 16 Blades, Fast Deployment & more...

iron SYSTEMS™
Iron Systems, Inc.
2330 Kruse Drive, San Jose, CA
www.ironsystems.com

CALL: 1-800-921-IRON

RESOURCES

"Customizing an eNote notebook with Debian" by Steve R. Hastings: www.linuxjournal.com/article/7165

Lindows MobilePC Information Page: info.lindows.com/mobilepc/mobilepc.htm

"Linux on eNote Travel Lite Notebook" by Andrew Comech: www.math.sunysb.edu/~comech/tools/enote-travellite

LETTERS CONTINUED FROM PAGE 6

Ahoy, Linux Cruisers!

I've just read Don Marti's review of Linux Lunacy 2003 on the Web [www.linuxjournal.com/article/7149]. It was great fun. Really worthwhile. For example, Guido van Rossum described the architectural decisions and trade-offs that are incorporated into Python. He tailored the talk to our interests and was happy to answer specific questions, even outside of the conference sessions. Don didn't mention the popular third stream with LPI training/certification. I'd also comment on the tremendous accessibility, inclusiveness and friendliness of the speakers and other attendees. There were plenty of technical discussions outside of formal lectures. Geek cruisers were really easy to spot hanging out in the Internet café, happily chatting and eating cake crowded together in the Lido, and burning the midnight oil in the Crow's Nest.

--
Alex Perry

Aboard Linux Lunacy 2003, *Linux Journal* authors from left to right: Alex Perry, Doc Searls, Don Marti, Greg Haerr (below) and Mick Bauer.

More on Desktop Linux, Please

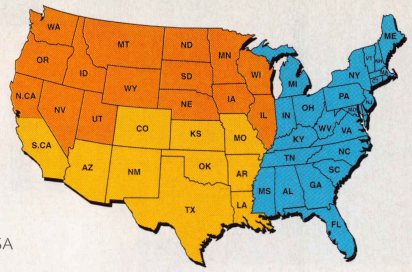
I read *Linux Journal* regularly and would really appreciate seeing your format expand to include a strong section devoted to the desktop—in particular, to reviewing, critically, applications of merit outside the traditional office suite and to providing installation instructions for a modest selection of currently available laptops that sell for about $1,300 or less. Years ago, I attempted to get Linux running productively in a desktop environment on a Dell laptop, and failed. I have been waiting to accomplish the same thing again (for over four years, to be exact), and every time I looked closely I found numerous unresolved issues with regard to laptops and power management in particular, font management and display (important to me), and no serious page layout applications.

--
Walter Hollenberg

We do cover individual laptop models on our Web site. There is a lot of new and exciting desktop software available, and we have a special desktop issue planned for March 2004.—Ed.

We welcome your letters. Please submit "Letters to the Editor" to ljeditor@ssc.com or SSC/Editorial, PO Box 55549, Seattle, WA 98155-0549 USA.

LINUX JOURNAL



PO Box 55549
Seattle, WA 98155-0549 USA
www.linuxjournal.com

ADVERTISING SERVICES

VP OF SALES AND MARKETING
Carlie Fairchild, carlie@ssc.com
+1 206-297-8655,
+1 206-782-7191 FAX

FOR GENERAL AD INQUIRIES
e-mail ads@ssc.com
or see www.linuxjournal.com/advertising

To receive further information from any of our advertisers, please visit our Reader Service page to access this service: **www.linuxjournal.com/advertising/rs**

REGIONAL ADVERTISING SALES

NORTHERN USA
Joseph Krack, joseph@ssc.com
866-423-7722 (toll-free),
866-423-7722 FAX

SOUTHERN USA
Annie Tiemann, annie@ssc.com
866-965-6646 (toll-free),
866-422-2027 FAX

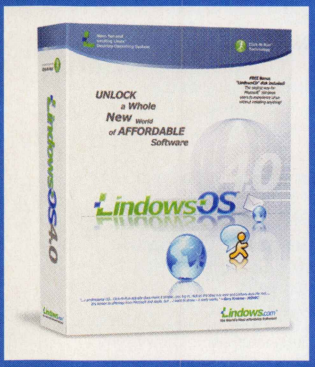
EASTERN USA AND CANADA
Martin Seto, mseto@ssc.com
+1 905-947-8846,
+1 905-947-8849 FAX

Please direct international advertising inquiries to VP of Sales and Marketing, Carlie Fairchild.

Advertiser	Page #
Angstrom Microsystems www.angstrom.com	1
Appro International, Inc. appro.com	43
Arkeia Corporation www.arkeia.com/demo	57
ASA Computers www.asacomputers.com	65, 85
Aspen Systems, Inc. www.aspsys.com	7
Central Command, Inc. www.centralcommand.com	73
Ciara Technology www.ciara-tech.com	17
Computer Associates (CA) www.ca.com/linux	21
Consensys Computers, Inc. www.consensys.com	C3
Coyote Point Systems www.coyotepoint.com	93
Cyclades Corporation www.cyclades.com	2, 23
Emac, Inc. www.emacinc.com	87
EmperorLinux www.emperorlinux.com	15
Etnus www.etnus.com	41
FairCom Corporation www.faircom.com	39
Free Software Foundation www.gnupress.org	58
Hentzenwerke www.hentzenwerke.com	27
Hurricane Electric www.he.net	47
ImageStream Internet Solutions www.imagestream-is.com	33
Integrated Computer Solutions, Inc. www.ics.com	69
InterSystems Caché www.intersystems.com	51
Iron Systems, Inc. www.ironsystems.com	80
Layer 42 Networks www.layer42.net	49
Linux Journal www.linuxjournal.com	53, 67, 72
LinuxWorld Conference & Expo www.linuxworldexpo.com	28, 29
Microway, Inc. www.microway.com	C4, 59
Mikro Tik www.routerboard.com	83
Paracel www.paracel.com	5
Penguin Computing www.penguincomputing.com	62, 63
Polywell Computers, Inc. www.polywell.com	55
The Portland Group www.pgroup.com	9
ProMicro Systems www.promicrosystems.com	37
RackSaver www.racksaver.com	C2, 19
Rackable Systems www.rackable.com	45
Rackspace Managed Hosting www.rackspace.com	11
Rae Internet, Inc. www.raeinternet.com	71
Sangoma Technologies, Inc. www.sangoma.com	61
SBE, Inc. www.sbei.com	31, 75
SGI www.sgi.com	25
SWsoft www.sw-soft.com	90
Tatung Science & Technology www.tsti.com	13
Technologic Systems www.embeddedx86.com	35

Lindows 4.0

REVIEWED BY STEVE R. HASTINGS



PRODUCT INFORMATION

Manufacturer:
Lindows.com, Inc.

URL:
www.lindows.com

Price:
$59.95 US (download version, $49.95 US)

THE GOOD

- Easy install with hardware detection.
- Slick package management.
- Family license might be a good deal.

THE BAD

- Poor security—defaults to always running as root.
- Rough edges could confuse new users.
- Many packages confusingly renamed.

Lindows is a Linux distribution intended to be as easy to use as possible, even for complete beginners. The Lindows promotional literature promised a ten-minute install, and indeed, the process took about ten minutes.

Lindows 4.0 comes with two CDs. One is the installation CD, which detects all your hardware and installs Lindows. The other is a demo CD, which detects all your hardware and then runs Lindows in RAM, loading any applications directly from the CD. You can use the demo CD to try out Lindows 4.0 on a computer without overwriting your data. I booted up the demo CD, and it worked as advertised. It let me experiment with Lindows and did not write to my hard disk.

Next, I installed Lindows on a spare computer. When you begin the installation you are offered two choices: automatic disk partitioning (labeled "take over the whole hard disk") or manual partitioning for advanced users. I let the Lindows installer take over the whole hard disk. It created a small boot partition and a 256MB swap partition; it then filled the rest of the disk with a single large ReiserFS partition.

The installer asked few questions, none of which concerned what kind of hardware the computer had; the installer detected hardware without prompting. At the end of the install, the installer prompted me to enter an optional system password. This password is, in fact, the root password. By default, Lindows sets up the user to run as root all the time and only weakly encourages the user even to set a password for root. This design certainly is a convenient way to set up a system, but it gives up much of the traditional security inherent in Linux.

Running

Lindows is based mostly on KDE. The desktop is a KDE desktop, heavily customized with the Lindows brand. The custom theme's icons run to green and blue colors that match the Lindows logo.

Lindows 4.0 defaults to Mozilla for both Web surfing and e-mail. Mozilla is rebranded as "Lindows Internet Suite", and unless you run `mozilla --version` at a shell prompt, it is difficult to figure out that it is version 1.3.

Konqueror, rebranded as File Manager and Web Browser, is the default file manager for Lindows. Desktop icons for My Computer, My Documents and Network Browser all open Konqueror to the appropriate location.

When I tried out a command prompt, I discovered that Lindows was missing some important tools, such as the **man** command. Fortunately, Lindows is based on Debian GNU/Linux and contains a working core Debian system. By running **apt-get** I was able to install man, vim and other must-haves for the more advanced user.

Click-N-Run

The most impressive part of Lindows is the Click-N-Run system. With Click-N-Run you can browse through a virtual warehouse divided into categories, with aisles (departments) containing software packages. For example, in the Multimedia & Design category you can find an aisle named Image Editing, which contains various image editors including The GIMP. If you click on the Click-N-Run icon next to The GIMP, it is added to your Click-N-Run queue and automatically installed.

Because Click-N-Run starts up as soon as you log in, it can download software continuously in the background while you work. As long as you are connected to the Internet, Click-N-Run downloads packages until your queue is empty again. You can open the Click-N-Run client to check on the status of your queue and to see what's already installed, see what's pending, cancel pending downloads and so on.

One nice touch, the Programs menu in the main program launcher is organized exactly the same as the Click-N-Run Warehouse. Thus, after you have installed The GIMP, you can find it by clicking on Start Applications, Programs, Multimedia & Design, The GIMP. The icons on the categories are the same icons used in the Click-N-Run Warehouse as well, so the system is visually well organized, too.

In order to use Click-N-Run, you need to sign up with Lindows for a Click-N-Run account. This costs from $50 to $150 US

per year, depending on what level of account you choose. Much of the Click-N-Run software is free software, such as The GIMP, but some proprietary software also is offered. If you use Click-N-Run to download a proprietary package, you are billed automatically.

Among the available proprietary software packages are VirusSafe, a virus scanner based on Vexira Antivirus, and SurfSafe, a Web site blocker based on the Cerberian Web Filter. With SurfSafe in place, trying to access a site like playboy.com brings up the "LindowsFamily — SurfSafe Warning" page.

Click-N-Run isn't perfect. When I installed MPlayer, it didn't work, but returned with no error message—it simply shut down. By running MPlayer from a command prompt, I was able to see an error message explaining that the .mplayer directory was not set up correctly. But Click-N-Run worked well for the other packages I tried.

I was somewhat annoyed at the relentless rebranding of packages that exists in Lindows 4. KWrite, the KDE word processor, is available in Click-N-Run, but it's called Write Pro. GTKPool is called Billiards; Gnumeric is rebranded as Numeric. Some other packages retain their names—but does a typical Lindows user know what TK Gocr is? If Lindows needs to come up with friendly names, I would hope they include the original name in the new one, as they do with *XGalaga Galactic Invaders*. Or, at least mention the original name somewhere on the Click-N-Run information page.

In the future, it should be possible to use Click-N-Run to upgrade from a previous version of Lindows to a newer one. However, Click-N-Run currently cannot upgrade from Lindows 3.0 to Lindows 4.0; you need to use the Lindows 4.0 CD.

Creating Users

At the end of Lindows setup, you have the option to create user accounts. Also, after setup, the standard KDE tool for managing users, KUser, is available in the Settings menu as User Manager. It's simple to create non-root users with this tool. Then, after the computer boots up, the KDM login manager prompts for a user name and password.

I highly recommend that you create a separate account for each person who will use the Lindows computer. For more on managing user accounts in Lindows, see my companion article on the *Linux Journal* Web site (www.linuxjournal.com/article/7166).

Family License

The Lindows Family License permits Lindows customers to install Lindows on multiple computers, as long as all the computers are in the customer's household. Click-N-Run has a section, My Products, that shows all software installed by the Click-N-Run user; this includes proprietary software licensed by the user. Any software the user has installed with Click-N-Run on a different computer is listed in a dim-grey font, while software installed on the current computer is shown in black. It's easy to click on grey packages and select them for downloading and installing on the current computer.

Thus, any proprietary software purchased using Click-N-Run actually is licensed for all Lindows computers in the purchaser's household. If you run proprietary software and your household has multiple computers, this license potentially could be a good deal.

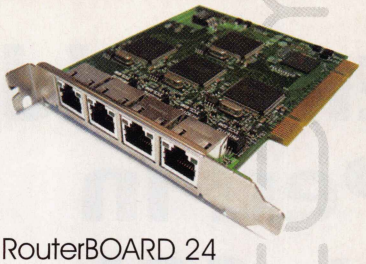
Conclusion

Click-N-Run is the slickest package installation system I have seen yet in a Linux distribution, but you must pay a minimum of $50 per year to use it. The Lindows system is mostly smooth and polished, but a few rough edges remain.

If you read *Linux Journal*, it is unlikely that you will want to run Lindows on your own computer. But, it is a possibility that family members or friends would appreciate a Linux system designed for newbies. As long as you set them up to not run as root all the time, Lindows is a good choice.

Steve R. Hastings first used UNIX on actual paper teletypes. He enjoys bicycling with his wife, listening to music, petting his cat and making his Linux computers do new things.

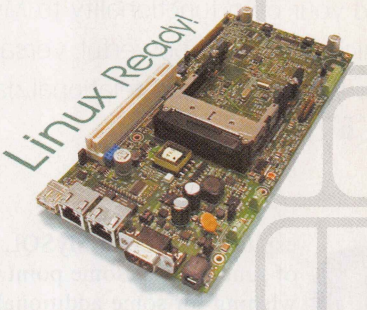
expand! $125



RouterBOARD 24

Four ethernet ports in one card that work as separate interfaces. Supported by all OSs!

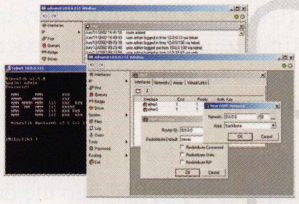
combine! $195



RouterBOARD 230

A super small integrated motherboard with various interfaces and 266 MHz CPU. Use as a CPE or an AP on a tower with 500ft of PoE (802.3af),or anything else you can imagine! Many cases and accessories available. IDE/CF, PCMCIA, miniPCI, PCI, USB.

route! $45



MikroTik.com RouterOS

A router operating system that has every possible feature a router could need - works with regular pc's and routerboards!

more!

Four-port miniPCI to PCI $95
One-port miniPCI to PCI $33
v.35 synchronous miniPCI $550
and much more on ...

RouterBOARD.com

sales@routerboard.com

RouterBOARD, RouterOS, and MikroTik are trademarks of Mikrotikls SIA

Embedding Perl in MySQL

Add your own functionality to MySQL with MyPerl, which brings the powerful, versatile Perl interpreter into the heart of the relational database.

BY BRIAN AKER

Despite the fact that MySQL comes with a rich set of functions, at some point you will find yourself wishing for some additional function or needing an advanced regular expression engine. To solve this problem, MySQL supports user-defined functions (UDFs). Through the UDF interface, you can load new functions into your database dynamically. Although this is a powerful feature, it does mean time spent debugging C or C++ code. As much as I like C, sometimes I don't have the time to write and debug applications written in it, and at other times I want a much faster development cycle. Enter Perl, the Swiss Army knife of languages, with a cast of thousands thanks to the Comprehensive Perl Archive Network (CPAN). Embedding Perl into MySQL gave me a lot of flexibility in being able to extend my database quickly. For these reasons, the embedded MySQL Perl interpreter, MyPerl, was written.

Setting Up Perl

The first step in putting Perl into the database is to get the right setup for Perl. Perl by default is not threadsafe, and MySQL, on the other hand, uses a thread for each user connection. So for Perl to live inside the database, you must compile a threadsafe version of Perl:

```
./Configure -Dusethreads -Duseithreads
```

Once this is finished and built, you will have a Perl that is threadsafe. This does not mean your code or any Perl modules you use will be threadsafe, it simply means that Perl itself will be. Building Perl to be threadsafe is a necessary step, because I know of no vendors shipping a threadsafe Perl at this time. Don't be fooled by the fact that MyPerl will build with a nonthreaded Perl; it can, but at some point it will crash your database. I suspect that with the advent of Apache 2.0 and a final release of mod_perl 2, some vendors will consider shipping their Perl binaries with threads enabled. While completing this article, I finally upgraded to Red Hat 9 and saw that they have begun shipping Perl with threads enabled.

Writing the Embedded Perl Interpreter

UDFs have three stages: init, request and deinit. The init stage is called once at the beginning of a query; the request stage is called once per row returned, and the deinit stage is called after the data is sent to the client. Both the init and deinit stages can be skipped, although for all but the most simple UDF you need to create and clean up memory you will use to return data to the client.

MyPerl starts off with the following init function:

```
my_bool
myperl_init(UDF_INIT *initid, UDF_ARGS *args,
            char *message)
{
    myperl_passable *pass = NULL;
    int exitstatus = 0;
    char *embedding[] = { "perl",
                          "-MMyPerl::Request",
                          "-e", "1" };
    PerlInterpreter *my_perl;
    uint i =0;
    initid->max_length = 256;
    initid->maybe_null=1;
```

Three parameters are passed into the init method, and it returns either success or failure. The UDF_INIT structure holds information that controls the behavior of how UDF responds. It also is the only structure that is passed between all three stages. First, MySQL is told the UDF will be sending data back that is greater than the size of a VARCHAR. By telling the server it needs more space then a VARCHAR, the server assumes it will be returning blobs.

Although MyPerl does not know that it actually will be doing this, at this point it has no way of knowing how much data it will be returning, so telling the server that it is returning blobs is the safer bet. Next, it sets maybe_null to 1, because there is always the possibility it will be returning NULL values. MyPerl returns NULL for both empty results and upon compilation errors that occur from code that you eval().

The next thing is to check the rows being passed in:

```
if (args->arg_count == 0 || i
    args->arg_type[0] != STRING_RESULT) {
    strncpy(message,USAGE, MYSQL_ERRMSG_SIZE);
    return 1;
}
for (i=0 ; i < args->arg_count; i++)
        args->arg_type[i]=STRING_RESULT;
```

MyPerl expects that the first row being passed in is the code it will execute. Therefore, if no rows are passed in or if the first row is not a string, an error should occur. Error messages must be at most the size of MYSQL_ERRMSG_SIZE, and they must be copied into the message string. To save some time, MyPerl walks through the arguments that MySQL will be passing in and tells it to convert them to strings.

Before the Perl interpreter is set up, a structure must be created to store the interpreter and track a memory block that it

Don't be fooled by the fact that MyPerl will build with a non-threaded Perl; it can, but at some point it will crash your database.

will use to return data for each request:

```
pass = (myperl_passable *)
    malloc(sizeof(myperl_passable));;
if (!pass) {
    strncpy(message, "Could not allocate memory",
            MYSQL_ERRMSG_SIZE);
    return 1;
}
```

The structure is as follows:

```
typedef struct {
    char *returnable;
    size_t  size;
    PerlInterpreter *myperl;
    size_t  messagesize;
} myperl_passable;
```

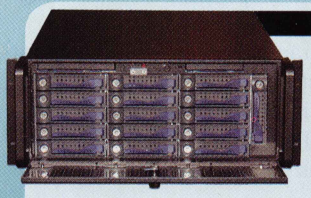
The char pointer returnable is used to store the block of memory, and size and messagesize are used to track both the total size and the current size of the returnable data. Because creating and destroying memory is quite costly, it is important to keep this down to a minimum. The Perl interpreter also will be stored in this structure.

At this point, the work of setting up the Perl interpreter that will be used for the query must be done. Currently, MyPerl creates a new Perl interpreter for each request to keep from leaking memory and ensure the security of data between requests. The odds of this becoming a pool of Perl interpreters in the future is quite high:

```
if((my_perl = perl_alloc()) == NULL) {
    strncpy(message, "Could not allocate perl",
            MYSQL_ERRMSG_SIZE);
    return 1;
}
perl_construct(my_perl);
exitstatus = perl_parse(my_perl, xs_init, 4,
                        embedding, environ);
PL_exit_flags |= PERL_EXIT_DESTRUCT_END;
if (exitstatus) {
    strncpy(message, "Error in creating perl parser",
            MYSQL_ERRMSG_SIZE);
    goto error;
}
exitstatus = perl_run(my_perl);
if (exitstatus) {
    strncpy(message, "Error in parsing your perl",
            MYSQL_ERRMSG_SIZE);
    goto error;
}
```


www.asacomputers.com

Hardware Systems for the Open Source Community—since 1989
(Linux, FreeBSD, NetBSD, OpenBSD, Solaris...)



4TB in 4U—$9,500
Intel E7500, 2 x Intel Xeon 2.0GHz
512 MB ECC DDR Max: 16GB
6 x PCI-X
2 x 10/100 LAN
1 Gig
Raid 0,1,5,etc.
CD, Floppy
Redundant Hot-Swap Power Supply

2U Dual Xeon SCSI—$2,899
6 Front-removable Hot-Swap SCSI bays
2 x Intel Xeon 2.8 GHz Max: 2
1 GB ECC DDR Max: 16GB
3 x 36GB U320 SCSI Drives
Raid 5 configured Max: 876GB Storage
CD, Floppy
Dual Gig ports
1 Extra 5.25 Drive Bay

Dual AMD Opteron in 1U—$2,749
2 x AMD Opteron 240
4 GB DDR ECC Registered Memory
1 x 40 GB IDE
Floppy, CD, Gig

AMD Opteron

Space-saving Mini-1U Dual Xeon IDE—$1,901
1U 14" Depth
533 MHz FSB
2 x Intel Xeon 2.8 GHz
1 GB DDR ECC Memory
40 GB IDE HDD
1 GIG, 1 10/100 LAN

1U P4 Entry-level Server—$735
1 x Intel P4 2.4 GHz
256 MB DDR ECC Memory
40 GB IDE Drive
CD, 2 x 10/100 LAN

1U Dual Xeon SCSI—$2,089
SuperMicro 6013P-8 Barebones
Dual Intel Xeon 2.8 GHz 533 FSB
512 MB DDR ECC Max: 12GB
18 GB 10K rpm SCSI
3 Hot-Swap SCSI Drive Bays
CD, Floppy
Dual Gig, VGA onboard

ASA Colocation
$50 per month for
1U Rack - 20 GB/month

ASA Colocation Special
First month of colocation free.*

Storage Solutions
IDE, SCSI, Fiber RAID solutions
TB storage options
3Ware, Promise, Adaptec,
ICP-Vortex, LSI, CMD, Chaparral,
Mylex controllers
JMR, Kingston/Storcase solutions

Clusters
Rackmount and Desktop nodes
HP, Intel, 3Com, Cisco switches
KVM or Cyclades Terminal Server
APC or Generic racks

All systems installed and tested with user's choice of Linux distribution (free). ASA Colocation—$50 per month

ASA Computers, Inc.
1-800-REAL-PCS
2354 Calle Del Mundo, Santa Clara, CA 95054
TEL: (408) 654-2900 • FAX: (408) 654-2910
E-MAIL: sales@asacomputers.com
Free delivery in Silicon Valley*
Open on Sat. too—10 to 1
Prices and availability of systems subject to change without notice
* Certain restrictions apply.

www.asacomputers.com
www.b2bforsale.com

The first function, perl_alloc(), allocates a new Perl interpreter that is then constructed with perl_construct(). Now it is simply a matter of starting up Perl. The embedding variable is used as parameters for the Perl interpreter. These are exactly the same parameters you would use on the command line. Errors must be checked for at every point in dealing with the Perl interpreter. In the current design, if an error occurs, MyPerl jumps to a collection of function calls that will clean up the memory that has been allocated.

Now that a good Perl interpreter exists, a few defaults in the pass structure need to be set; the interpreter must be stored, and the address of the structure must be stored in initid->ptr pointer so it can be used throughout the query:

```
pass->returnable = NULL;
pass->size = 0;
pass->messagesize = 0;
pass->myperl = my_perl;
initid->ptr = (char*)pass;
return 0;
```

After all of this setup, MyPerl is ready to start taking requests:

```
char *
myperl(UDF_INIT *initid, UDF_ARGS *args,
       char *result, unsigned long *length,
       char *is_null, char *error)
{
    myperl_passable *pass =
        (myperl_passable *)initid->ptr ;
    char *returnable = NULL;
    unsigned long x = 0;
    size_t size = 0;
    char *newspot = NULL;
    char *string = NULL;
    myperl_passable *pass =
        (myperl_passable *)initid->ptr ;
    STRLEN n_a; //Return strings length
    PerlInterpreter *my_perl = pass->myperl;
```

It is important that the address of the interpreter is copied into a variable named my_perl. Much of the Perl internals are based on macros that expect you to use variables with certain names. STRLEN is a variable type that Perl uses to store the size of strings.

At this point the interpreter is called:

```
dSP;
ENTER;
SAVETMPS;
PUSHMARK(SP);
// Now we push the additional values into ARGV
for(x = 0; x < args->arg_count  ; x++) {
    XPUSHs(sv_2mortal(newSVpvn(args->args[x],
           args->lengths[x])));
}
PUTBACK;
call_pv("MyPerl::Request::handler", G_SCALAR);
SPAGAIN;
string = POPpx;
size = (size_t)n_a;
```

XPUSHs is used to push all of the rows into an array of strings that will be passed to the Perl function handler() in the library MyPerl::Request(). This Perl module is similar to the Apache::Request module except where the Apache module uses a filename to track code it has eval'ed, MyPerl uses the code itself to determine this.

Because the variable named size now holds the size of the data that will be returned, space needs to be allocated for it:

```
if (size) {
    if(pass->size < size) {
        newspot = (char *)realloc(pass->returnable,
                                  size);
        if(!newspot) {
            error[0] = '1';
            returnable =  NULL;
            goto error;
        }
        pass->size = size;
        pass->returnable = newspot;
    }
    // Always know the current size,
    // it may be less than the full size
    pass->messagesize = size;
    memcpy(pass->returnable, string, size);
} else {
    is_null[0] = '1';
}
error:
PUTBACK;
```

w3m: w3m.sourceforge.net

Three years ago I reviewed several good programs, but the hands-down winner for me was the text-based browser w3m. With the large majority of systems today using frames and images, using a text browser on the Internet can be a frustrating experience. Although great for slow links or systems not running X, this is not ideal. However, w3m works well with frames, allowing you to see multiple frames on the screen. It also, depending on your build options, allows you to see some graphics. The older Lynx has long since been replaced on my system with w3m. Requires: libgc, libpthread, libm, libnsl, libgpm, libncurses, libssl, libcrypt o, glibc, libstdc++, libdl and libgcc_s.

—DAVID A. BANDEL

```
FREETMPS;
LEAVE;
*length = pass->messagesize;

return pass->returnable;
}
```

This is where the information that needs to be sent to the server is stored. The memory call realloc() is used if more memory needs to be allocated. If no data is received from the Perl interpreter, is_null is set to 1 so that MySQL knows a null result should be returned to the client. MyPerl also makes sure to clean up memory that might have been used for the call_pv() function.

The myperl() function is now called for each row. After MySQL returns its data to the client, it calls the deinit function to free the interpreter and release any allocated memory:

```
void myperl_deinit(UDF_INIT *initid)
{
    myperl_passable *pass =
        (myperl_passable *)initid->ptr ;
    perl_destruct(pass->myperl);
    perl_free(pass->myperl);
    free(pass->returnable);
    free(initid->ptr);
}
```

Examples Using MyPerl

Now that a Perl UDF exists, simple tricks like this can be performed:

```
mysql> select myperl(´return $ARGV[0]´,User)
       from mysql.user;
```

As you can see, each row corresponds with a value in @ARGV. You also can use MyPerl with CPAN modules to enter data directly. This example fetches the content for a list of URLs and inserts the content into the database:

```
mysql> insert into html select
       myperl("use LWP::Simple;
       my $content = get($ARGV[0]);
       return $content", url) from urls;
```

Using modules like XML::Simple and XML::XPath, you can even query any XML you may have stored in your database. I have used MyPerl for quickly debugging serialized Perl objects I have stored in the database.

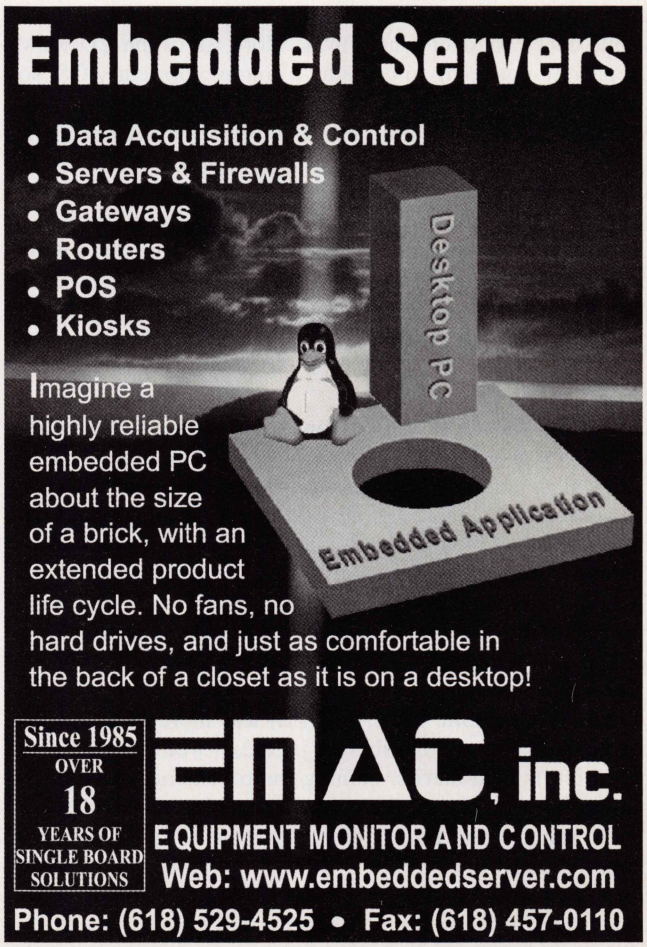
But What about GROUP BY?

Although the above demonstrates how to handle row requests with this code, it does not work for queries that use GROUP BY to treat data as sets. For this reason, there is an additional type of UDF called an aggregate. Aggregates differ from their more bland cousins by having two additional stages, reset and add. With aggregate UDFs, the add function handles each row, and the request stage sorts out the results and sends the data on to the client. The reset stage is called at the beginning of each data set, so it is guaranteed to be called at least once. MyPerl currently has an aggregate UDF, but its design is still in flux.

Conclusion

By embedding Perl into MySQL, the realm of possibilities as to what you can do in your database has expanded. Although frequently it is better to keep your database as simple as possible, you may find in some cases this is not practical. Imagine having to pull a gigabyte of text out of the database and send it to a client to be used. The time spent on sending the data would be considerable; being able to do work on the data directly in the database with Perl may turn out to be a real time-saver. Being able to make use of Perl advanced regular expressions may allow you to write simple clients in other languages that do not have a good regular expression support. I am sure you will find many uses for this in your own environments. MyPerl can be found at software.tangent.org along with other UDFs you can use as examples to write your own.

Brian Aker (brian@tangent.org) spends his time working on MySQL and Apache modules, which include mod_layout and the Apache streaming services module, mod_mp3. He recently coauthored the *Running Weblogs with Slash* book for O'Reilly. For years he worked on Slashdot and now works for MySQL AB as a senior software architect. He also teaches the Perl Certification course at the University of Washington. He lives in Seattle, Washington with his dog Rosalynd.



Cross-Platform CD Index

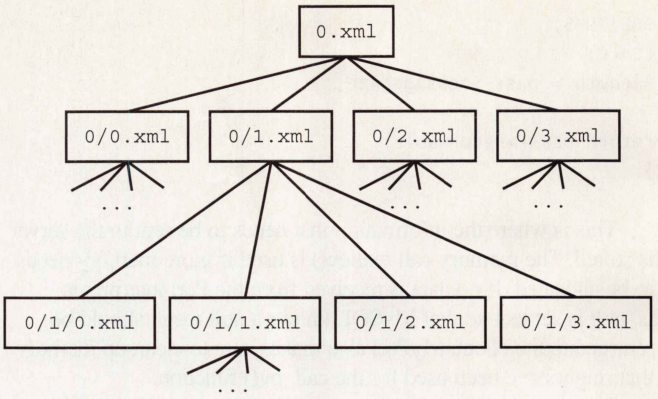


Figure 1. jsFind creates a B-tree, where an XML file represents each node.

CD-ROM content needs a search engine that can run in any browser, straight from static index files. JavaScript and XML make it possible.

BY SHAWN P. GARBETT

I recently was working on a CD-ROM catalog for a client, and he requested that it have keyword search ability. My searches for solutions to such a request kept turning up proprietary OS software that required an install on the user's machine and a license fee paid per copy distributed. Such installation requirements are limiting and would cost a lot over time. Furthermore, all of the CD-ROM users were not going to be using a single proprietary OS, so this automatically reduced the potential customer base. While sitting back to think about the situation, a package in my mailbox caught my eye—the *Linux Journal* Archive CD. I figured if anybody had solved this problem, it was sure to be on the *LJ* Archive. Imagine my disappointment upon discovering that the *LJ* Archive CD has a really good index but no search engine. If a solution was to be found, I would have to find it myself. This article is about scratching that proverbial itch with jsFind.

Licensing

One of my earliest considerations was how to distribute and license my solution, jsFind. I showed early versions of it to colleagues, and they felt I should follow the model in which I license the code and then market it. jsFind then would be using the same model as other competing search engines for this type of content. Personally, I would rather spend my time coding than marketing, and I suspect the total market is not very large. I would rather get informative CD-ROMs and be able to search them easily using any browser and platform I choose.

The GNU Public License (GPL) was more in line with my goals. By freely distributing jsFind, it would be marketed based on its own merits, gaining improvements and contributions as it grows. At the risk of preaching to the choir, one of the goals of proprietary systems is to lock users in to being required to use their system by every possible means. For example, when one gets a CD-ROM and is required to use a specific browser and a specific OS to use the search engine, then that user is forced to access a copy of that OS. CD-ROM producers also are forced to keep buying development tools for that OS in order to stay current. The result is consumers and producers are locked in to the proprietary OS vendor. Releasing jsFind under the GPL would break the cycle.

How It Was Done

The jsFind keyword search engine itself is a small JavaScript program of about 500 lines. A browser that supports DOM Level 3 JavaScript extensions can load XML files. The current versions of Mozilla, Netscape and Microsoft Internet Explorer all support these extensions, and the upcoming release of Konqueror will do so as well. The index is stored as a set of XML files, and the JavaScript searches through these in an efficient manner to generate results of a keyword search. These results then can be posted back to the Web page that requested them, also using JavaScript.

One of the key dependencies of jsFind is that a CD-ROM be a set of static information. Unlike Web search engines or any other dynamic data set, once pressed, a CD-ROM isn't going to change. SWISH-E is better suited for dynamic indexing, especially when one has the luxury of configuring a server to do the keyword searches. Therefore, jsFind is based on the idea that the only thing available is a standard Web browser with JavaScript and a set of browseable files—a severe restriction on possible solutions.

Most indexing method algorithms try to strike a balance between insert, update, delete and select times. Because a CD-ROM is static, there will never be a delete or update. Insert takes place prior to CD burning and can be quite time consuming. Select time is critical for user responsiveness. An additional constraint of small space is required, because a typical CD-ROM can't hold more than 700MB.

Re-examining indexing methods based on these constraints yielded an interesting solution: B-trees and hashes are the two most commonly used indexing methods. I chose to use B-trees due to the fact that a filesystem organizes files in a tree; this could be used to store the structure of the B-tree, saving some precious space in the process. Second, the key/link pairs could be analyzed, and a bal-

anced B-tree could be created. The structure of the XML files themselves was kept as minimal as possible, so single-letter tags were used as a space-saving move.

Description of B-trees

A B-tree is a data structure used frequently in database indexing and storage routines. It offers efficient search times, and storage/retrieval is done in blocks that works well with current hardware. A B-tree consists of nodes (or blocks) that have an ordered list of keys. Each key references an associated data set. If a requested key falls between two keys in the ordering, a reference is provided to another node of keys. A balanced B-tree is one in which the maximum number of nodes that could be loaded on a search stays at a minimum.

jsFind creates a B-tree by using XML files for the nodes of the tree, and the directories on the filesystem correspond to references to another set of nodes. This allows for part of the structure of the B-tree to be encoded on the filesystem. If all the XML files are in the same directory, file open times might become long, so using the filesystem efficiently requires subdirectories.

Short Tutorial

End users need not worry about any of this. They simply can type words to search for on a Web page, and jsFind returns links to pages containing those keywords. No install, no worries, just a seamless experience.

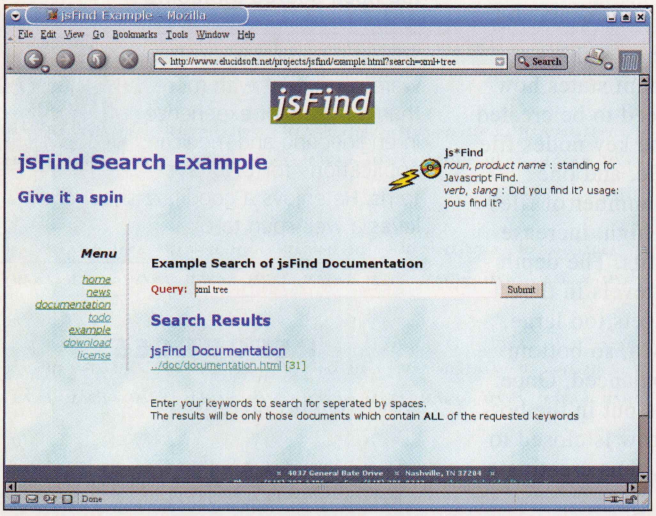


Figure 2. Example Search Results

As a developer of content, however, your life is not so simple. The jsFind toolset tries to make your job as easy as possible, though. To start, you need Perl and a fair amount of CPU time to generate the index. Most likely you also need a copy of all the target browsers so you can test the results. An example with a Makefile can be found in the jsFind distribution, but several steps need to be tailored to your individual needs.

The first step is to get a data set consisting of keywords and links; the input format needs to be XML. I used SWISH-E with a custom patch to extract and create an index and then exported the results to the XML format suitable for processing with jsFind's Perl scripts. Assuming the SWISH-E index is in the file mystuff.index, the following command exports the file to XML:

```
$ swish-e -f mystuff.index -T INDEX_XML > mystuff.xml
```

The structure of this XML file is as follows:

```
<index>
 <word>
  <name>akeywordhere</name>
  <path freq="11" title="Something neat">
    /cdrom/blah.html
  </path>
  <path freq="10" title="More cool stuff">
    /cdrom/blah2.html
 </path>
 </word>
 <word>
 ...

</index>
```

The XML file is sorted by order of keyword name.

The resulting data set still is probably too large, because SWISH-E doesn't concern itself with filtering out words like "and", "this" and other common English words. Two Perl programs can be used to filter the result, occurrences.pl and filter.pl. occurrences.pl creates a list of keywords and determines the number of times they occur in an index:

```
$ occurrences.pl mystuff.xml | sort -n -k 2 \
  > mystuff.keys
```

This file has a keyword on each line followed by the number of occurrences:

```
$ tail mystuff.keys
you 134910
for 138811
i 149471
in 168657
is 179815
of 252424
and 273283
a 299319
to 349069
the 646262
```

At this point, the mind-numbing task of creating a keyword exclusion file is performed. Edit the key file and leave in all the words that should be excluded from the final index. Even better than creating your own file, get a copy of the 300 most common words in English from ZingMan at www.zingman.com/commonWords.html.

Next, run the filter. The Perl script filter.pl included in this package filters a result set. It currently is set to exclude any single-character index keys (except the letter C), any key that starts with two numeric digits (so things

PLESK 6
#1
CONTROL
PANEL
SOFTWARE
IN THE WORLD
SWsoft
www.sw-soft.com
sales@sw-soft.com
+1 (703) 815-5670

like 3com and 0xe3 are okay) and anything in the specified exclusion file:

```
$ filter.pl mystuff.xml mystuff.keys > \
  mystuff-filtered.xml
```

This step takes quite a bit of time. Make sure the final size of the file falls within the limits of the space available. The final index should be about 75% of the size of the filtered index. If it's too big, whittle it down to size with a longer keyword exclusion file.

The second big step is creating the index itself. A script is provided to break this index down into a set of B-tree XML files:

```
$ mkindex.pl mystuff-filtered.xml 25
blocksize: 20
keycount: 101958
depth: 4
blockcount: 5098
maximum keys: 194480
fill ratio: 0.524259563965446
bottom fill: 92698
Working: 11%
```

Parameters are the next thing to consider. The blockcount states how many B-tree blocks need to be created. Each block creates one key nodes file and one data nodes file, and one directory. If the total number of files and directories is too high, increase the blocksize until it fits. The depth shows the number of levels in the tree. If the blocksize gets too large, search times slow down, so bottom fill is how it is kept balanced. Once that number of keys is put in the bottom row, the bottom row is closed to further node creation, thus creating a balanced tree.

If all works well, you should end up with three files in the current directory: 0.xml, _0.xml and the directory 0. These are the index files. The next step is to follow the provided example for integrating the results into your HTML/JavaScript. The results then are passed to a provided routine and need to be posted back to the current Web page. The example does this using JavaScript to create dynamic HTML.

Conclusion

Many improvements to jsFind are possible, and they'll come as it is used by open-source users. Such features as having an image archive search with thumbnails, multipage result sets and stronger browser compatibility checks all are possible using this code as a springboard.

The 2002 *LJ* Archive CD-ROM contains the jsFind search engine. If you're a developer of CD-ROM content, please consider using jsFind over solutions with proprietary OSes. Doing so will be cheaper and will connect you with a larger potential user base. As an end user, I hope to be delivered from having to install a program and dual-boot simply to search content on a CD-ROM. Other innovative uses for the software might be possible as well, so consider it to be one more tool in the open-source toolbox.

Shawn P. Garbett is a software consultant with more than 15 years of experience in engineering and medical applications for UNIX systems. He enjoys a good jazz show and loves a well-spun tale.

RESOURCES

GNU: www.gnu.org

"How to Index Anything" by Josh Rabinowitz, *LJ*, July 2003: www.linuxjournal.com/article/6652

jsFind is available at: www.elucidsoft.net/projects/jsfind

TECH TIP

Type `make help` in the top-level directory of the Linux 2.6 source code for a list of helpful targets you can make. If you're building a small kernel for your specific hardware, try `make allnoconfig` to get a stripped-down configuration with everything turned off, then use `make menuconfig` to go back in and add the drivers and features you need.

DVD Transcoding with Linux Metacomputing

A Condor high-throughput DVD transcoding system for Linux.

BY F. J. GONZALEZ-CASTAÑO, R. ASOREY-CACHEDA, R. P. MARTINEZ-ALVAREZ, E. COMESAÑA-SEIJO AND J. VALES-ALONSO

As a consequence of the many recent advances in video and audio encoding, the MPEG-2 format now is used for digital video broadcasting (DVB) transmission and DVD storage and is supported by a wide range of hardware devices. MPEG-2 movie files typically range in size from 3–6GB, sizes that are suitable for DVDs but not for CD-Rs. Similarly, high-quality MPEG-2 videos are suitable for DVB-S or DVB-T networks, but not for IEEE 802.11b or domestic HomePlug transmission. To solve these kinds of problems, improved encoding techniques have been developed, and as a result, MPEG-4 has been standardized. The MPEG-4 format can reduce movie sizes down to 700MB or so and maintain reasonably good quality.

Because much multimedia content is available as DVD MPEG-2 files, it is necessary to transcode them to obtain the MPEG-4 equivalents. In this article, we propose a Linux framework based on the Condor metacomputing platform to achieve high-throughput DVD transcoding. Although some LAN parallel transcoding tools for fixed sets of machines exist, we are not aware of any metacomputing system for parallel transcoding. Metacomputing refers to architectures that hide physical resources and instead offer a simplified virtual machine view. For example, the Condor tool we use "steals" cycles of available machines when neither users nor high-priority processes are using them.

Background

The DVD movie market has boomed thanks to the availability of cheap DVD players, the robustness of DVD as a storage media as compared to VHS cassettes and so on. The DVD recording media market, however, is incipient. Because CD-R technology has been around for a while and CD-R disks are much cheaper than DVD disks, domestic users have found ways to store DVD movies on CDs with similar subjective qualities. This kind of storage is possible due to the last generation of video and audio codecs. They are based on the MPEG-4 standard and offer high compression ratios. Transcoding a DVD to make its contents fit in a CD, however, still is expensive computationally for many desktop PCs.

Parallelization is a promising solution to accelerate DVD transcoding. The most obvious approach is manual parallelization, dividing input files in chunks manually, transcoding the chunks in different machines and joining the result in a single file. Manual parallelization may be adequate for users who wish to keep track of the whole process. However, it may be advantageous to use metacomputing to implement high-throughput, submit-and-forget DVD transcoding.

Parallelizing a process requires breaking it into elementary tasks, scheduling those tasks and collecting their results. Consequently, a resource management tool is necessary. Tools such as Condor and Globus provide basic metacomputing and parallelization software. In our case, we have chosen Condor because it does not add extra complexity, it is easy to install and configure and it works properly on Linux. Finally, Condor does not require a dedicated cluster.

Condor is a specialized workload management system for computation-intense jobs. Like other full-featured batch systems, Condor provides a job queuing mechanism, a scheduling policy, a priority scheme, resource monitoring and resource management. Users submit their serial or parallel jobs to Condor, and Condor places them into a queue, chooses when and where to run the jobs based on a policy, monitors their progress and ultimately informs the user of a job's completion.

While providing functionality similar to that of any traditional batch queuing system, Condor's architecture allows it to succeed in areas where traditional scheduling systems fail. Unique mechanisms enable Condor to harness wasted CPU power from otherwise idle desktop computers. For instance, Condor can be configured to use desktop machines only when the keyboard and mouse are idle. Should Condor detect that a machine is no longer available (say, a key press is detected), it is able to produce a transparent checkpoint and migrate a job to a different machine that would otherwise be idle. Condor also is able to redirect transparently all the job's I/O requests back to the submitting machine. As a result, Condor can be used to combine seamlessly all the computational power in a community.

The apparent lack of commercial metacomputing transcoding systems may exist because metacomputing mostly has been linked with the UNIX scientific community. On the other hand, entertainment software designers still give maximum priority to the metacomputing-unfriendly Microsoft Windows world. For example, the most recent version of the DivX codec—v.5.0.5 at the time this article was written—is a key tool for Linux transcoding development, but it did not work properly on Pentium 4 Linux boxes. The previous release was v.5.0.1alpha, an unstable version that had been released the previous year. This example provides an idea of the problems one may encounter when trying to port entertainment applications to metacomputing-friendly Linux platforms.

Although diverse transcoding applications are available, we outline the three that we found most interesting:

- FlaskMpeg: one of the first transcoding applications to appear. Currently, it is one of the most popular in the

Windows world. It does not support parallelization.

- Mencoder: one of the top Linux applications for DVD transcoding. Its efficiency (output-to-input size ratio) in general, is slightly worse than FlaskMpeg's. As in the previous case, it does not support parallelization.

- Dvd::rip: a high-level Linux transcoder based on another program, Transcode. Its results are comparable to those of Mencoder. Dvd::rip does support parallelization, but it is difficult to configure. Parallelization requires manual configuration of all computers involved in the transcoding process. This configuration is static, and it does not react to environmental changes (a major difference for a Condor-oriented system like ours). Dvd::rip does not admit audio streams. The audio stream must be processed sequentially due to technical problems Dvd::rip points out but does not solve; see Dvd::rip's Web page. This is a minor problem, though, because transcoding time is dominated by video transcoding, regardless of whether the audio transcoding strategy is employed in parallel or sequentially.

DVD Partitioning

DVD is based on a subset of standards ISO/IEC 11172 (MPEG-1) and ISO/IEC 13818 (MPEG-2). A DVD movie is divided into three parts: video objects (VOBs) files with a maximum size of 1GB each, multiplexing video and audio sources.

Three types of MPEG-2 frames exist: I (Intra), P (Predictive) and B (Bidirectionally-predictive). I frames represent full images, while P and B frames encode differences between previous and/or future frames. In principle, it seems obvious that video stream cuts must be located at the beginning of I frames. This is almost right, but not quite. Some parameters, such as frame rate and size, must be taken into account. This information is part of the Sequence Header. For this reason, packets chosen as cut points must have a Sequence Header. Fortunately, there is a Sequence Header before every I frame.

Another important issue is frame reordering due to the existence of P and B frames. After an I frame, B frames may follow that depend on P frames that came prior to the I frame. If the video stream is partitioned at the start of that I frame, it is not possible to maintain video transcoding consistency. The solution consists of assigning the late B frames to the previous chunk. As a consequence, a little extra complexity is added to video preprocessing.

Obviously, it is not interesting to fragment video to the maximum extent, because the size of the chunks would be too small. Typically, about 300KB exist between two consecutive I frames, although this length depends on several parameters, such as bit rate or image size.

Load Balancing

We considered two basic load balancing strategies for our project. In the first, called Small-Chunks, the DVD movie is divided into small chunks of a fixed size. Condor assigns a chunk to every available computer. When a computer finishes transcoding one chunk, it requests another one. This process is repeated until there are no more chunks left on the server. In the other strategy, called Master-Worker, load balancing depends on the shares, which are determined by the master processor. Obviously, the other computers involved are the workers. This strategy often is used for high-throughput computations. For this project, chunk size for each particular computer is assigned according to a training stage, as explained in the next section.

It should be understood that we deliberately do not consider the possibility of machine failures or user interference. If those events take place, the performance of the simple Master-Worker implementation in this project would drop. Nevertheless, our two approaches are illustrative because they are extreme cases, pure Master-Worker on one hand and the high granularity of Small-Chunks on the other. Fault/interference-tolerant Master-Worker strategies lie in the middle. Our aim is to evaluate whether the behavior of our application is similar in the two extremes, in terms of processing time and transcoded file size. As the results described in this article suggest, Small-Chunks may be more advantageous due to its simplicity (it does not need a training stage) and because it adapts naturally to Condor's management of machine unavailability.

Master-Worker Training Stage

To provide information to the Master-Worker coordinator, it is necessary to evaluate all computers beforehand. Evaluation is performed in a training stage, which estimates the transcoding rate of each computer in frames per second. The training stage of our prototype consists of transcoding a variety of small video sequences in the target computer set and estimating the average frames per second delivered by each computer. This result then is used to set the sizes of the data chunks, which are proportional to the estimated performance of each computer. Ideally, this approach minimizes DVD transcoding time, because all computers should finish their jobs at the same time.

Test Bed Layout

Our test bed emulated a typical heterogeneous computing environment, including machines at the end of their usage lives. It was composed of five computers (see Table 1), classified in three groups according to their processing capabilities. Two machines were in the first group (gigabyte and kilobyte), a single computer was in the second group (nazgul) and two machines with the worst performance (titan and brio) were in the third group.

Table 1. Test Bed Computers

Name	CPU	MHz	Memory	Kflops	Mips
gigabyte	Intel Pentium 4	1,700	256 DDR	528,205	1,388
kilobyte	Intel Pentium 4	1,700	256 DDR	624,242	1,355
nazgul	Intel Celeron	433	192	152,593	491
titan	Intel Pentium II	350	320	67,987	398
brio	Intel Pentium II	350	192	72,281	398

Load balancing	Comp	Video transcod. *t*	Movie montage *t*	Total time	Fps
Small-Chunks	g----	1:51:14	0:01:32	2:03:52	19.6
	gk---	0:58:25	0:02:57	1:09:00	35.1
	gk--b	0:53:53	0:03:32	1:04:31	37.6
	g-ntb	1:13:04	0:03:54	1:24:55	28.6
	gkntb	0:47:33	0:04:07	0:58:55	41.2
Master-Worker	g----	1:49:11	0:01:31	1:59:23	20.3
	gk---	0:56:23	0:03:03	1:06:10	36.7
	gk--b	0:51:06	0:02:37	1:00:06	40.3
	g-ntb	1:08:00	0:03:00	1:17:38	31.2
	gkntb	0:42:55	0:03:58	0:52:35	46.1

Table 2. Computational Results (t = time, Fps = frames per second)

In addition, all computers were linked to a 100Mbps Ethernet network, and the operating system used in all computers was Red Hat Linux 8.0. All computers shared the same user space, defined by an NIS server, and the same filesystem (NFS server in gigabyte). Finally, we installed Condor v. 6.4.7, and gigabyte was the central manager. Condor was configured to keep all jobs in their respective processors regardless of user activity. Thus, the timings in this section are best-case results, as mentioned above.

The DVD-to-DivX parallel transcoder was implemented with the following libraries:

- libmpeg2 0.3.1: DVD MPEG-2 stream demultiplexing and decoding.
- liba52 0.7.5-cvs: DVD AC3 audio decoding.
- DivX 5.0.1alpha: MPEG-4 video encoding.
- lame 3.93.1: MP3 audio encoding.

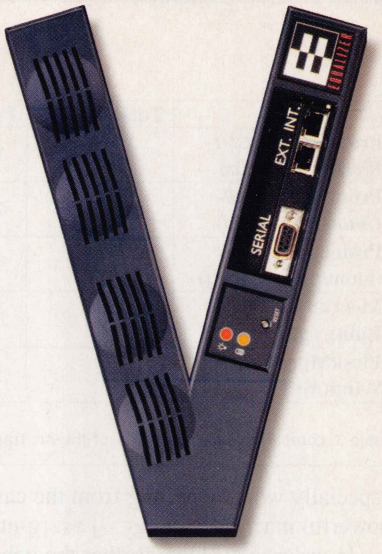
Parallel Video Transcoding

Once the video partitioning stage is done, video data chunks are submitted to Condor transcoding jobs. These jobs are processed in the Condor Vanilla universe, because they load the DivX library dynamically.

In order to transcode a data chunk, every transcoder reads the data directly from source VOBs. Output data is written to the same folder. Read/write operations are performed on a frame-to-frame basis: a transcoder reads a frame, transcodes it and writes the result back. This strategy yields a better performance than delivering whole chunks to the workers, transcoding them in worker local filesystems and sending whole transcoded results from the workers to the server computer for joining. All computers used NFS to share both input VOB files and the output folder. Once the parallel transcoding stage finishes, transcoded results are a set of independent files, which are concatenated at the master to generate a DivX movie. Table 2 presents the results.

We used the two load balancing strategies, Small-Chunks and Master-Worker. The testing movie was *All about My Mother*, which has a length of 1 hour and 37 minutes and an original size of 2.94GB. The tuples in the comp column in Table 2 are the first letters of the names of the test bed computers, for example, g refers to gigabyte and t refers to titan. A - symbol indicates that the computer was not used in that particular test. Chunk size in Small-Chunks was set to 60MB. Video preprocessing time has not been included because it was negligible in all cases.

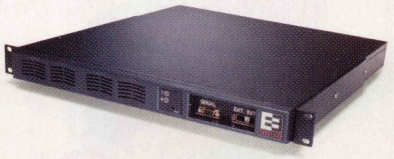
Several conclusions can be extracted from Table 2. First, according to the Fps column, load balancing is better with Master-Worker than it is with Small-Chunks. The difference is small, but it tends to grow as the number of machines used increases. In general, parallelization increases transcoding performance, which is evident when adding a second powerful machine (see [g----] vs. [gk---]). The impact of adding low-end machines successively is low (see [gk---] vs. [gk--b] and [gk--b] vs. [gkntb]). However, the combined impact of all low-end machines is noticeable,



Win the battle with server costs.

(Load balancing: the affordable way to attack server capacity limits.)

If traffic is jamming your servers, the answer isn't more servers. It's more traffic capacity. Get it with the Equalizer, the affordable load balancer that distributes requests between servers here, there or anywhere on your network. So end the assault on your IT budget. Call **650-969-6000** today or visit **www.coyotepoint.com**



COYOTE POINT SYSTEMS INC

@2002 Coyote Point Systems, Inc.

	Total Time	Fps	AVI size (MB)
Prototype, *Master-Worker* (gkntb)	0:52:35	46.1	739.2
Prototype, *Small-Chunks* (gkntb)	0:58:55	41.2	742.8
Prototype (monoprocessor, g)	1:59:23	20.3	739.1
Mencoder 0.90 Linux (g)	2:01:02	20.0	731.1
FlaskMpeg 0.78.39 Windows (g)	1:53:12	21.4	724.1

Table 3. Comparative Results, State-of-the-Art Transcoding Applications

especially when departing from the case of a single available powerful machine (see [g----] vs. [g-ntb]).

In order to evaluate further the behavior of the prototype, we compared it with two popular transcoding tools, Mencoder and FlaskMpeg. Table 3 shows these results. The speed of the monoprocessor version of our prototype lies between FlaskMpeg's and Mencoder's. Regarding output size, in the worst case (Small-Chunks), our prototype delivers a DivX movie that is only 2.6% larger than FlaskMpeg's output. Indeed, the global compression rates achieved by Small-Chunks (24.67%) and FlaskMpeg (24.05%) are similar, and the difference is not relevant if processing speedup is taken into consideration. It is important to note that FlaskMpeg uses DivX codec v. 5.0.5 Pro, which was not available for Linux at the time this article was written. Therefore, compression performances may be even closer when the Linux version becomes available.

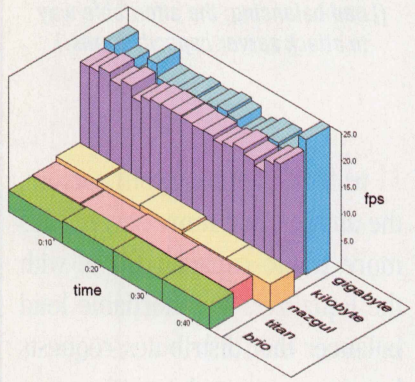


Figure 1. Individual Computer Throughput, Small-Chunks

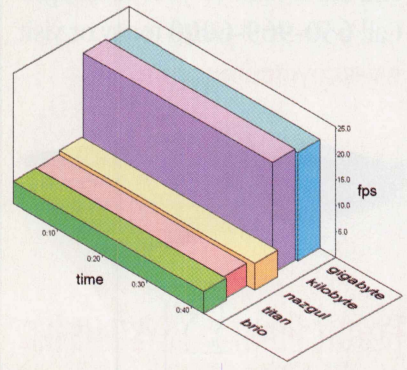


Figure 2. Individual Computer Throughput, Master-Worker

Finally, Figures 1 and 2 show the throughput of the machines in the system prototypes, both Small-Chunks and Master-Worker, for load balancing. The computers do not finish their assignments exactly at the same time. This should be expected, though, as Small-Chunks load balancing is only approximate. Plus, Master-Worker job size is assigned according to the results of a training stage, which is representative but not exact.

Conclusion

In this article, we have presented a Condor high-throughput DVD transcoding system for Linux. Our results indicate that metacomputing-oriented parallel transcoding is of practical interest, and it can achieve noticeable improvements when compared to existing monoprocessor Windows tools.

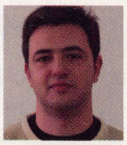
Attending to the statistics of our case study, pure Master-Worker produces better results than Small-Chunks, but the difference is minimal and seems irrelevant in practice.

Francisco J. Gonzalez-Castaño is currently a Professor Titular with the Departamento de Ingenieria Telematica, Universidad de Vigo, Spain and has been a visiting assistant professor with the Computer Sciences Department, University of Wisconsin-Madison. He is the head of the TC-1 Information Technology Group, Universidad de Vigo. His research interests include mobile communications, high-performance switching, metacomputing and data mining.

Rafael Asorey Cacheda was born in Vigo, Spain in 1977. Currently, he is a researcher with the TC-1 Information Technology Group, University of Vigo. His interests include content distribution, high-performance switching, video transcoding and IPv6.

Rafael P. Martinez-Alvarez works as a telecommunications engineer for the TC-1 Information Technology Group, University of Vigo, Spain. His interests include multimedia encoding formats and real-time multimedia transcoding.

Eduardo Comesaña-Seijo was born in 1976 in Vigo, Spain. He has been a researcher with the TC-1 Information Technology Group, University of Vigo. He currently works for Comunitel Global SA (a Spanish Telco). His interests include real-time multimedia transcoding and parallel computing.



Javier Vales-Alonso is a Professor Ayudante with the Department of Information Technologies and Communications, Polytechnic University of Cartagena, Spain. His research interests include mobile networks and metacomputing.

RESOURCES

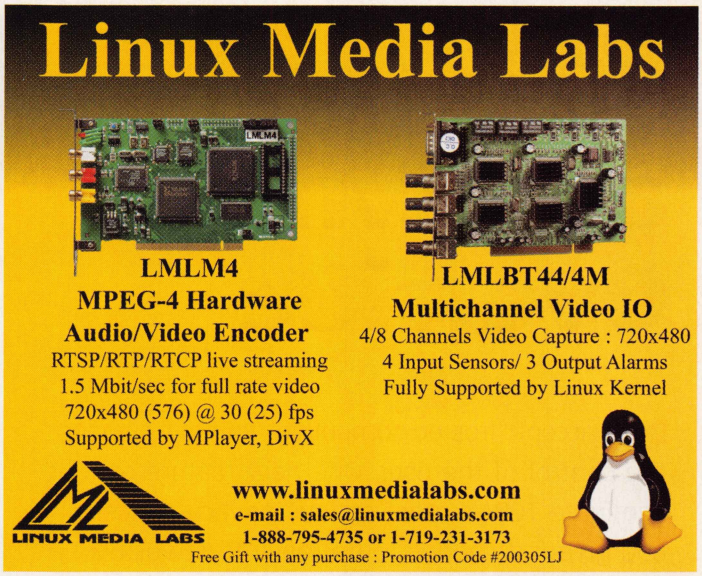
Condor Project:
www.cs.wisc.edu/condor

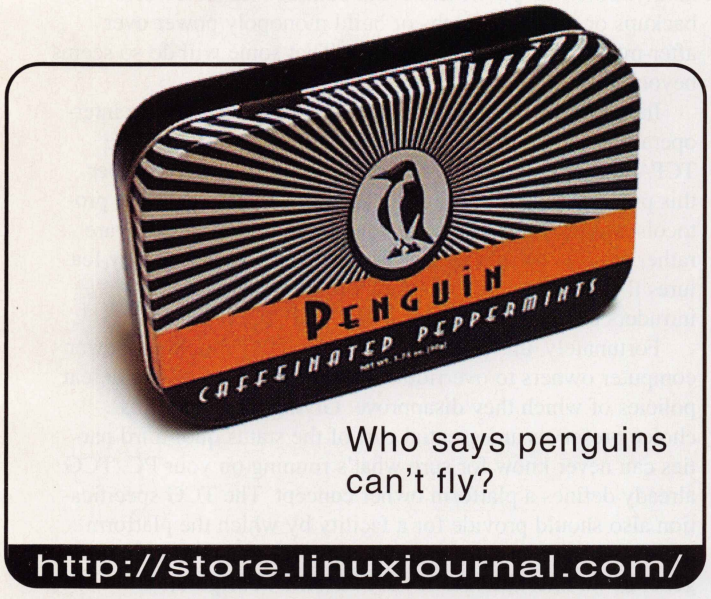
Dvd::rip:
www.exit1.org/dvdrip

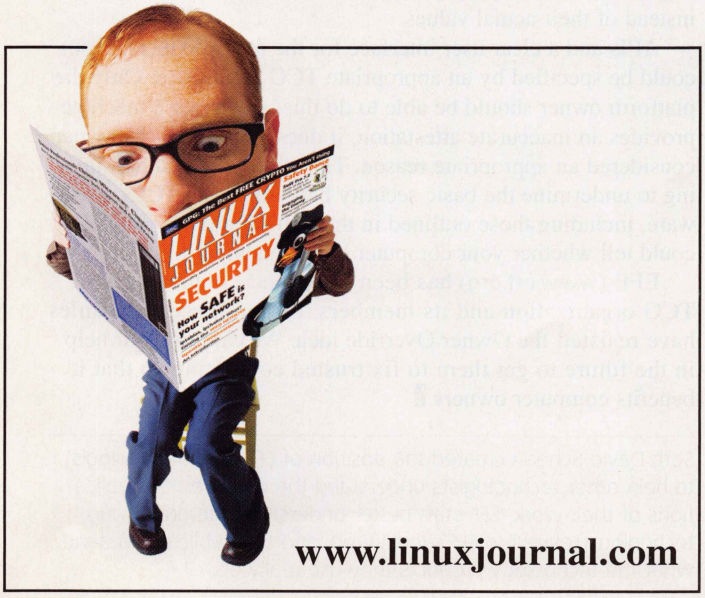
Globus Project:
www.globus.org

Mencoder:
www.mplayerhq.hu/DOCS/encoding.html

Transcode:
www.theorie.physik.uni-goettingen.de/~ostreich/transcode







FREE NEWSLETTER!

Wish you could get the latest from *LJ* more than once a month? You can—sign up today for ***LJ*'s weekly e-mail newsletter**.

Each week the *LJ* newsletter features great tech tips, links to web-only articles, and news on the latest events in the Linux market.

Sign up for the *LJ* e-mail newsletter now:

http://www.linuxjournal.com/

Give TCPA an Owner Override

Don't accept "trusted computing" unless it's under the control of the user. One new feature can eliminate TCPA's most serious privacy concern and give it a chance to be really trustworthy.

BY SETH DAVID SCHOEN

The Trusted Computing Platform Alliance (TCPA) overview in *Linux Journal*'s August 2003 issue, "Take Control of TCPA" written by Dave Safford, Jeff Kravitz and Leendert van Doorn, discusses some of the TCPA system's security features. The article does not consider, however, that the remote attestation feature of TCPA is likely to cause harm unless a simple fix is applied.

TCPA's basic rationale is correct: today's PC architecture lacks hardware features that would improve computer security. For instance, current PCs have no secure key storage mechanism. If you keep your GPG key on your computer, as most GPG users do, someone who breaks in is able to read it from disk and copy it.

The PC also is vulnerable to keyloggers and screen-grabbers. An attacker can watch you by installing code to record what you type and can capture what appears on your monitor.

TCPA improves computer security by addressing these weaknesses with corresponding hardware improvements. One of these, the secure key storage mechanism described in "Take Control of TCPA", lets you prevent keys from being stolen, even in the case of a root compromise.

The Trusted Computing Group (TCG) also may develop hardware to defend against software keyloggers and screen-grabbers and to support stronger memory isolation than what exists in most popular operating systems today. Even if your system is compromised, your data and communications still could be protected. This approach would help programmers limit the privilege of software, even operating systems—that's good security design.

But trusted computing architects have gone astray in designing "system software integrity measurement", which Safford et al. note "can be used to detect software compromise". The TCPA software attestation mechanisms go beyond this; they're built to enforce policies even against the wishes of the computer owner.

Even as the attestation features prove *to you* that your computer's software is unaltered, they also can be used to prove *to a third party* that your computer's software is unaltered. If you, as the computer owner, make a change, a third party sees that the content of your attestation is different. That's fine if the third party has your interests at heart, but computer owners and third parties with whom they interact often have different interests.

Creating a reliable way for a third party to determine what software you're using is a pernicious project. Today, it's trivial to fool "Internet Explorer only" sites: change how your browser identifies itself, and there is nothing the other end can do. With TCPA's remote attestation, a site that insists on an attestation would receive the whole truth or a highly suspicious "No Comment".

Another casualty might be the software interoperability that we currently take for granted. For example, today you can access Microsoft file servers with a Samba client. With a software attestation scheme that treats you as an adversary, software publishers and service providers would have a remarkably sophisticated toolset for locking out competition, forcing upgrades or downgrades and deliberately breaking interoperability. Software publishers could control their customers' backups or migration paths or build monopoly power over after-market products and services. That some will do so seems beyond dispute.

In the past, anticonsumer misfeatures and barriers to interoperability were remedied by reverse engineering. But the TCPA security design allows software publishers to hamper this practice severely. They can create encrypted network protocols and file formats—with keys tucked away in hardware rather than accessible in memory. Thus, the same security features that protect your computer and its software against intruders can be turned against you.

Fortunately, this problem is fixable. TCG should empower computer owners to override attestations deliberately to defeat policies of which they disapprove. Giving the owner this choice preserves an essential part of the status quo: third parties can never know for sure what's running on your PC. TCG already defines a platform owner concept. The TCG specification also should provide for a facility by which the platform owner, when physically present, can force the TPM chip to generate an attestation as if the Platform Configuration Registers (PCRs) contained values of the owner's choice instead of their actual values.

APIs and a clear user interface for the override mechanism could be specified by an appropriate TCG committee. Only the platform owner should be able to do this; whenever a machine provides an inaccurate attestation, it does so for what its owner considered an appropriate reason. This change would do nothing to undermine the basic security benefits of the TCPA hardware, including those outlined in the Safford article; *you* still could tell whether your computer had been altered.

EFF (www.eff.org) has been in regular contact with the TCG organization and its members. So far, these companies have resisted the Owner Override idea. We can use your help in the future to get them to fix trusted computing so that it benefits computer owners.

Seth David Schoen created the position of EFF Staff Technologist to help other technologists understand the civil liberties implications of their work, EFF staff better understand the underlying technology related to EFF's legal work and the public understand what the technology products they use really do.

New RAIDZONE *WireSpeed*

Unmatched Performance in All Network Storage Architectures --- NAS, SAN, iSCSI

New Products

Gangstor: Replicated Storage Made Easy

With disk capacities doubling and prices falling, you can buy two RAIDZONE servers for last year's price of one. The importance of data replication and business continuation makes that a very good idea. Our new Gangstor replicated servers make it a no-brainer.

Heard About SAN & NAS Converging? With AllStor, They've Mated

RAIDZONE AllStor combines NAS (NFS, WIndows), iSCSI, and Fibre Channel SAN in one easy-to-configure server appliance. Just create partitions, and allocate them individually to any storage protocol.

OpenNAS for SAN: The Ultimate Front End NAS

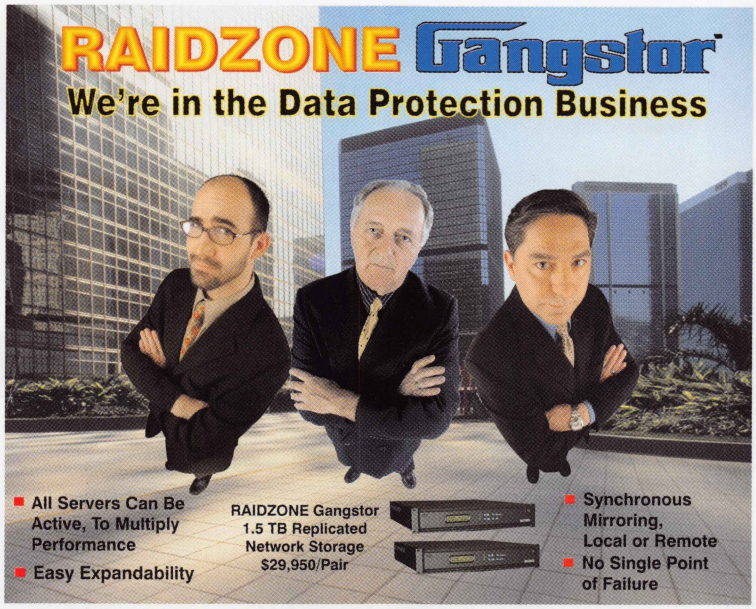
You've invested in a high end SAN. Now get the fastest, most reliable NAS front end. With up to 3 TB of redundant local storage as a cache/mirror of your allocated SAN storage, RAIDZONE OpenNAS for SAN effortlessly serves all your NFS and SMB (Windows) clients and provides transparent replication of your data.

For more information, please visit our Web site:

www.raidzone.com

or give us a call at:

1-800-388-1896



The *WireSpeed* RAIDZONE OpenNAS

NFS and Windows (SMB) At Over 100 MB/Sec.

RS8 2U 1.4 TB $14,450

RS15 4U 3 TB $24,950

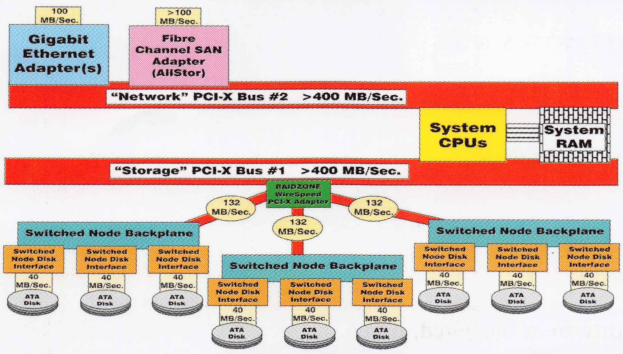
The *WireSpeed* RAIDZONE AllStor

SAN & NAS in One Server: iSCSI, Fibre Channel, NFS, SMB

RA15 4U 3 TB $26,950

RA8 2U 1.4 TB $16,450

RAIDZONE *WireSpeed* Server Architecture



New Ideas

Get a Free 21-Day Eval

Our network storage products sell themselves. If you've got an immediate requirement, we'll send out one of our best salesmen --- AllStor, Gangstor, or OpenNas for San --- to win your business.

RAIDZONE WireSpeed Data Path: Two 500 MB/Sec. PCI-X Buses

Our RAIDZONE disk array technology was the first to connect every disk independently to a 133 MB/Sec. PCI system bus. The new RAIDZONE WireSpeed technology uses one 500 MB/Sec. PCI-X system bus for the disk array, and a separate 500 MB/Sec. PCI-X bus for the network interface.

RAIDZONE WireSpeed Servers Are Serial ATA-Ready

We pioneered the use of Ultra ATA disk arrays in enterprise class storage. Since mid-2000, hundreds of thousands of RAIDZONE NAS users have reaped the benefits in bigger, faster network storage at much lower prices. Today, on the verge of a new generation of ATA drives, we're still leading. Our WireSpeed Servers can use both Ultra ATA and Serial ATA drives.

RAIDZONE is a registered trademark of Consensys Corp.; OpenNAS is a registered trademark and Gangstor, AllStor are trademarks of Consensys Computers Inc.; Windows/Microsoft Corp.; Linux/Linus Torvalds. The image of any identifiable person in this advertisement is used for illustrative purposes only; no endorsement is implied.

Technology you can count on

64-bit!

Itanium® 2 and Opteron™ Clusters

Robust Solutions for Demanding HPC Applications

Itanium2, Opteron, Xeon, Athlon and Alpha Clusters

Expert Integration...

Microway offers innovative, competitively priced, custom designed clusters with NodeWatch™ management and monitoring tools. Our solutions incorporate the latest processors, proprietary cooling and storage solutions, plus high-speed Myrinet and InfiniBand interconnects for both fine and coarse grain parallel applications.

Superior Service and Tech Support...

We understand that on-time delivery, out-of-the-box reliability and excellent ongoing technical support are critical to our users. Microway offers professional services from specialists with a wide range of expertise in HPC applications. On-site installations and training are also available.

Satisfied Customers...

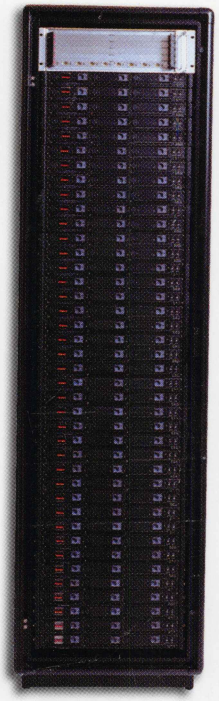
AT&T, Cessna, GE, GSK, Johnson & Johnson, LANL, LLNL, MBL, Millennium Pharmaceuticals, NIH, Northrop Grumman, Raytheon, Sandia, Seagate, US Air Force, Army, Navy, NASA, NOAA and hundreds of leading universities are among our satisfied customers since 1982.

"The Brain Imaging Research Center (a joint center of Carnegie Mellon University and University of Pittsburgh) decided to purchase our Linux cluster from Microway because of the proven performance of their clusters at Carnegie Mellon in the processing of high-volume brain imaging data. Microway was flexible and helpful at all stages, starting from the initial custom configuration and ending with timely delivery and full installation."

– Marcel Just, Co-Director, Brain Imaging Research Center

Call us first at ***508-746-7341***
for quotations and benchmarking services.
Find technical information, testimonials, and online newsletter at ***www.microway.com.***

See us at **SC2003**, Booth 2227
Phoenix – November 17-20



CoolRak™ Cabinet
Features 3 10" 535 CFM Fans, yielding 7 degrees cooler ambient per node. Available only on Microway 1U Opteron, Athlon and Xeon chassis.

Microway 1U CoolRak™
features dual 2.8GHz Xeons with proprietary Across the Board™ Cooling

Innovative Technology and Services at Competitive Prices

- **Xeon XBlade™ Technology by Microway** Incorporates two dual Xeon XBlades in our 1U chassis for 880 Gigaflops performance per 44U cabinet. Each Xeon XBlade provides two 133 MHz PCI-X sockets enabling next generation interconnects including InfiniBand.
- **NodeWatch™** provides automatic remote monitoring of vital cluster parameters and failsafe shutdown of the cluster. NodeWatch monitors temperatures, voltages, and chassis fans, runs off the master, and can be controlled by a secure web-based GUI. Available on Microway 1U Opteron, Xeon and Athlon chassis.
- **Integrated solutions powered by Platform Computing** Workload Management and Grid Computing software solutions to optimize computing resources. From Microway's Clusterware-powered CW Series, to integration of LSF 5 on any Microway cluster, all Platform solutions are "grid ready" for future system enhancements.
- **Integrated professional MPI/Pro®** enhances cluster value with middleware for high performance, scalable and robust message passing.
- **Fully redundant, highly-available storage systems** based on 2Gbit fiber channel technology for multi-terabyte storage requirements. State-of-the-art storage directors for full access to storage from any cluster node. Easily scales to address rapidly expanding storage requirements.
- **Consulting Services:** Microway offers expertise in designing custom software for financial analysis, genomics, proteomics, plus data warehousing, project management tools, network/data center design, and lab integration/management systems.

Microway®
Technology you can count on℠

AMD Opteron

intel inside ITANIUM 2

intel inside XEON